Hierarchical Spatiotemporal Aspect-Based Sentiment Analysis for Chain Restaurants using Machine Learning

Mouyassir Kawtar, Abderrahmane Fathi, Noureddine Assad, Ali Kartit
National School of Applied Sciences, The Information Technology Laboratory at Chouaib Doukkali University
El Jadida, Morocco

***Abstract*—In recent years, aspect-based sentiment analysis of restaurant business reviews has emerged as a pivotal area of research in natural language processing (NLP), aiming to provide detailed analytical methods benefiting both consumers and industry professionals. This study introduces a novel approach, Hierarchical Spatiotemporal Aspect-Based Sentiment Analysis (HISABSA), which combines lexicon-based methods such as VADER Lexicon, the AFFIN model, and TextBlob with contextual methods. By integrating advanced machine learning (ML) techniques, this hybrid methodology facilitates sentiment analysis, empowering chain restaurants to assess changes in sentiments towards specific aspects of their services across different branches and over time. Leveraging transformer-based models such as RoBERTa and BERT, this approach achieves effective sentiment classification and aspect extraction from text reviews. The results demonstrate the reliability of extracting valid aspects from online reviews of specific branches, offering valuable insights to business owners striving to succeed in competitive markets.**

Keywords*—HISABSA; hybrid model; NLP; ML; VADER Lexicon; AFFIN model; TextBlob; ABSA; Restaurant reviews; Transformer-based models; Lexicon-based methods; RoBERTa model; BERT model

I. Introduction

Understanding the sentiments of customers is essential for perceiving their emotional connections with a place or business of personal significance, often derived through the analysis of customer reviews [1]. The solicitation of customer opinions not only stimulates a sense of connection between customers and the organization but also serves as a key tool in evaluating how well a service or product aligns with customer expectations. This valuable feedback provides businesses with insights that can influence decision-making processes, contributing to increased profitability or reduced marketing expenses [2]. In the domain of selecting a new dining restaurant, various factors, including food quality, service, staff interactions, and facilities, come into play. The expansion of online reviews, particularly on platforms like Google Reviews, TripAdvisor [3], and Yelp [4], has replaced old-fashioned advertising as a primary influence on customer decision-making [5].

With the rise of artificial intelligence (AI), businesses can now automate the reading and analysis of reviews, eliminating the need for manual intermediation and reducing costs [6]. Advanced machine learning and deep learning within the domain of Natural Language Processing (NLP), have proven instrumental in the semantic analysis of these reviews. In addition, the use of Aspect-Based Sentiment Analysis (ABSA) allows a detailed examination of multiple dimensions within customer feedback, offering valuable insights into their preferences and priorities [7].

NLP plays an important role in ABSA for restaurant reviews, employing advanced machine learning and linguistic algorithms to break down complex, unstructured textual data into meaningful components [8]. This method enables the extraction of specific aspects such as food quality, service, ambiance, and others, providing a comprehensive understanding of customer sentiments [9]. By delving into different aspects in a text review, NLP helps identify positive and negative sentiments, facilitating targeted improvements to enhance the overall dining experience. ABSA, operating within the NLP framework, delves into linguistic patterns, lexical resources, and syntactic structures for its rule-based methodologies [10].

Machine learning (ML) and deep learning (DL) algorithms necessitate labeled or annotated data for model training. ML models involve manual feature engineering techniques to refine the extracted features for training [11]. In contrast, DL models eliminate the need for feature engineering, autonomously extracting significant features. The neural network learns these features through parameter adjustments and error calculation during training [12]. Consequently, the development of advanced DL models becomes essential for business intelligence and decision-making based on sentiment analysis of customer reviews.

In this exploration of sentiment analysis methodologies, this study go beyond the usual methods, incorporating three distinct lexicon-based methodologies: VADER, AFINN, and TextBlob, each contributing unique algorithmic strengths to the comprehensive analysis of customer sentiments. Furthermore, this research extends into the field of advanced models, such as BERT and RoBERTa, indicating a big change in sentiment analysis. BERT, known for its contextualized embeddings, and RoBERTa, a robust transformer-based model, highlight the significant impact of these models in revealing detailed insights from customer reviews [13].

This article navigates through the evolving landscape of sentiment analysis methodologies, highlighting the transformative impact of AI in extracting detailed insights from customer reviews. This exploration extends beyond the usual sentiment analysis to reveal the multifaceted dimensions of customer sentiments, providing businesses with a comprehensive overview for strategic improvements and informed decision-making.

II. Literature Review

Aspect-Based Sentiment Analysis has emerged as a pivotal research domain within the broader scope of NLP and sentiment analysis. This projection is expended by the exponential growth of user-generated content on diverse online platforms. In recent years, ABSA has experienced notable advancements, particularly with the integration of rule-based methods in the NLP domain [14]. Researchers have proposed various approaches for sentiment analysis that leverage predefined linguistic rules based on opinion lexicons, syntactic and semantic indications, such as parts-of-speech (POS) tags and lexical indicators [15]. POS tags serve to identify the grammatical category of each word, while lexical indications service in recognizing sentiment expressions. These particularly constructed rules are designed to capture specific linguistic patterns, facilitating the identification of sentiments associated with different aspects of the text.

Rule-based methods, known for comprehensibility and controllability in the sentiment classification process, have found popularity among researchers [16]. However, these approaches encounter challenges when confronted with complex linguistic variations and contextual differences [17]. Efficiently capturing linguistic indicates and adapting to evolving language usage necessitate the incorporation of advanced techniques to address these limitations [18]. Researchers in the domain are actively involved in formulating strategies to proficiently classify various textual aspects and their associated sentiments [19].

Working with restaurant reviews on platforms such as Yelp, TripAdvisor, and Google Reviews, and understanding the different sentiments expressed within these reviews has become increasingly important for both consumers and businesses [20]. This literature review aims to delve into the extensive body of research dedicated to ABSA, specifically focusing on restaurant reviews from different datasets.

In recent years, machine learning techniques for ABSA have seen widespread application, surpassing the performance of rule-based methods remaining to their efficient extraction of features and context information. In a study by the authors [21], a Naïve Bayes (NB) classifier employing chi-square (Chi2) for feature selection achieved an F1-score of 78.12%. Another notable investigation [22] devised an architecture for ABSA utilizing Convolutional Neural Network (CNN) and bidirectional Recurrent Neural Network (Bi-RNN) to capture local features and semantic information. Subsequently, support vector machine (SVM) was applied for classifying sentiments as positive or negative towards aspect terms. The evaluation on a French smartphone dataset and the SemEval-2016 restaurant dataset revealed an F1-scores of 94.05% and 85.70%, respectively. Notably, this combined architecture demonstrated superior performance compared to individual Bi-RNN, CNN, and SVM models.

Wang et al. [23] explored various approaches for ABSA, including a hierarchical bidirectional long short-term memory (Bi-LSTM), word-level attention model, clause-level attention models, and word & clause-level attention models. The findings indicated that the last approach, word & clause-level attention, outperformed all others with an F1-score of 0.68 and 0.66 for restaurant and laptop datasets, respectively.

Several recent studies focus on categorizing distinct aspects in a sentence but overlook the interference caused by other aspects within that sentence. Addressing this concern, [24] introduced a model known as Multi-Aspect-Specific Position Attention Bidirectional Long Short-Term Memory (MAPA BiLSTM) combined with Bidirectional Encoder Representations from Transformers (BERT). This model underwent evaluation with different datasets, achieving an F1-score of 85.73% for Multi-Aspect Multi-Sentiment (MAMS), 80.78%, 87.33% for SemEval2014 (restaurant and laptop reviews), and 75.31% for the Twitter dataset.

In a different study of Khan et al. [25], social media review data was employed for Aspect Category Detection (ACD). Employing a combination of next-word prediction, next-sequence prediction, and pattern prediction techniques, the researchers developed a convolutional attention-based bidirectional modified LSTM for ACD. The model underwent evaluation using state-of-the-art datasets, achieving an F1-score of 78.96% for SemEval-2015, 79.10% for SemEval-2016, and 79.03% on the SentiHood dataset. Another approach proposed by [26] introduced an IAN-BERT (Interactive Attention Network-BERT) model for determining the sentiment orientation of aspects in sentences. This model utilized a Post-trained BERT trained on Yelp and Amazon datasets, deviating from the generalized BERT trained on Wikipedia and BookCorpus datasets.

In a separate study leveraging the YELP dataset, researchers focused on improving restaurant businesses through sentiment analysis [27]. Various machine learning, deep learning and transfer learning approaches were employed, resulting in notable achievements: Logistic regression with an F1-score of 83.72%, NB with 73.35%, CNN with 87.85%, BERT (pre-trained) with 89.47%, and ALBERT (pre-trained) with 89.21%.

The authors of the research [29] compared various techniques and models for sentiment analysis of Yelp reviews. Liu explored a range of ML techniques such as SVM, Naive Bayes, Gradient Boosting (XGBoost), and Random Forests, alongside DL models including Recurrent Neural Networks (RNNs) and CNNs. By comparing the performance of these techniques and models, Liu aimed to identify the most effective approach for sentiment analysis in the context of Yelp reviews.

A research by Boya Yu et al. [30] addresses the challenge of evaluating restaurant quality beyond overall ratings on platforms like Yelp. Recognizing the need for more detailed evaluations involving aspects such as environment, service, and flavor, the study introduces a machine learning-based method to discern these features for specific restaurant types. The primary approach involves employing a support vector machine (SVM) model to analyze the sentiment expressed in each review based on word frequency. A summarized overview of existing approaches for sentiment analysis of online reviews is presented in Table I.

TABLE I. STATE-OF-THE-ART RESEARCH IN SENTIMENT ANALYSIS

Ref	***Dataset***	***Technique***	***Performance (f1-score)***	***Remarks***
[21]	SemEval 2014 Task 4 about product reviews	Chi-square and NB	78.12%	Enhancements with larger datasets and different MLmethods
[22]	8,000 French smartphone reviews (Amazon) SemEval-2016 restaurant dataset	CBRS (CNN-Bi-RNN-SVM)	Smartph dataset: 94.05% SemEval-2016: 85.70%	Further improved using transfer learning
[23]	SemEval-2015 (Laptop and restaurant datasets)	Hierarchical Bi-LSTM Word-Level ATT Clause-Level ATT Word&Clause-Level ATT	Restaurant 0.647% 0.662% 0.659% 0.685% For laptop 0.632% 0.646% 0.647% 0.667%	The performance can be improved using relationships between different clauses
[24]	SemEval2014 (Laptop and restaurant review datasets), Twitter review, and multi aspect multi-sentiment (MAMS) dataset	MAPA BiLSTM)	SemEval: 80.78% Restaurant: 87.33% Twitter: 75.31% MAMS: 85.73%	Improvement using ablation experiments
[25]	SemEval-2015 (2885 reviews), SemEval-2016 (3041 reviews), SentiHood (5215 sentences)	Convolutional attention-based bidirectional modified LSTM	78.96% 79.10% 79.03%	Enhancement using aspect category to improve customer satisfaction
[26]	SemEval-2014 (Restaurant and Laptop), MAMS	IAN-BERT	Restaurant: 81.5% Laptop: 80.3% MAMS: 83.2%	Improving performance by applying graph attention instead of BERT.
[27]	YELP dataset	LR NB CNN BERT ALBERT	83.72% 73.35% 87.85% 89.47% 89.21%	Improvements using embeddings of a specific domain and attention mechanism.
[28]	YELP	SVM RF Multinomial NB KNN	0.76% 0.78% 0.77% 0.61%	Utilizing more robust feature extraction techniques.
[30]	YELP	SVM	Accuracy: 88%	Based on a high-accuracy SVM model, calculating word scores

The researchers of [31] leveraged ML algorithms such as SVM, Naive Bayes, and Random Forests to classify sentiments based on word frequency and context. Additionally, the researchers explored DL methodologies, including RNNs and Convolutional Neural Networks (CNNs), which excel at capturing complex patterns in textual data. By combining these ML and DL approaches, Hemalatha et al. aimed to extract detailed sentiments from Yelp reviews, enabling a comprehensive understanding of customer opinions and preferences regarding various businesses and services.

III. ARCHITECTURE OF THE HYBRID SENTIMENT ANALYSIS MODEL

A. Overview of the Architecture

The central aim of this research is to develop a tool that reveals valuable insights in evaluating the sustained performance of a business over an extended period, targeting specific periods of time and spatial aspects based on the locations of its branches. This contribution involves introducing a thorough and all-encompassing approach designed to provide a comprehensive evaluation of branch-specific restaurants.

The YELP dataset serves as a source for experimentation and sentiment analysis in this study, covering 150,346 businesses, 11 metropolitan areas, and 6,990,280 reviews in JSON format. A subset of data containing restaurant reviews categorized by state was extracted, forming the groundwork for a multi-aspect sentiment analysis. Branch-specific data was derived from the state-wise dataset, enabling both Lexicon-Based sentiment analysis and Aspect-Based sentiment analysis. This spatiotemporal analysis is designed to closely examine individual branches within a specific restaurant business, revealing expressed sentiments by users across various aspects, including food, service, environment, and more.

This research attempts to construct an advanced hybrid sentiment analysis model, as illustrated in Fig. 1. This model integrates two distinct approaches: lexicon-based sentiment analysis and context sentiment analysis, with the aim of achieving a complete comprehension of sentiment dynamics. In the lexicon-based segment of sentiment analysis, the objective of this study is to compute a standardized star rating derived from customer reviews using models such as VADER, AFINN, and TextBlob. This process helps reduce inaccuracies caused by differences between written reviews and customer star ratings, ensuring a more equitable evaluation.

Furthermore, in the context sentiment analysis segment, this methodology introduces an enhanced technique that relies on extracted star ratings to classify customer reviews. BERT-based methods are utilized to extract aspects from textual reviews and sort them, enabling a deeper exploration of customer sentiments across different dimensions. By integrating these methodologies, the hybrid model aims to provide a robust framework for sentiment analysis, offering to the diverse traces of customer feedback and developing the efficacy of sentiment classification processes.

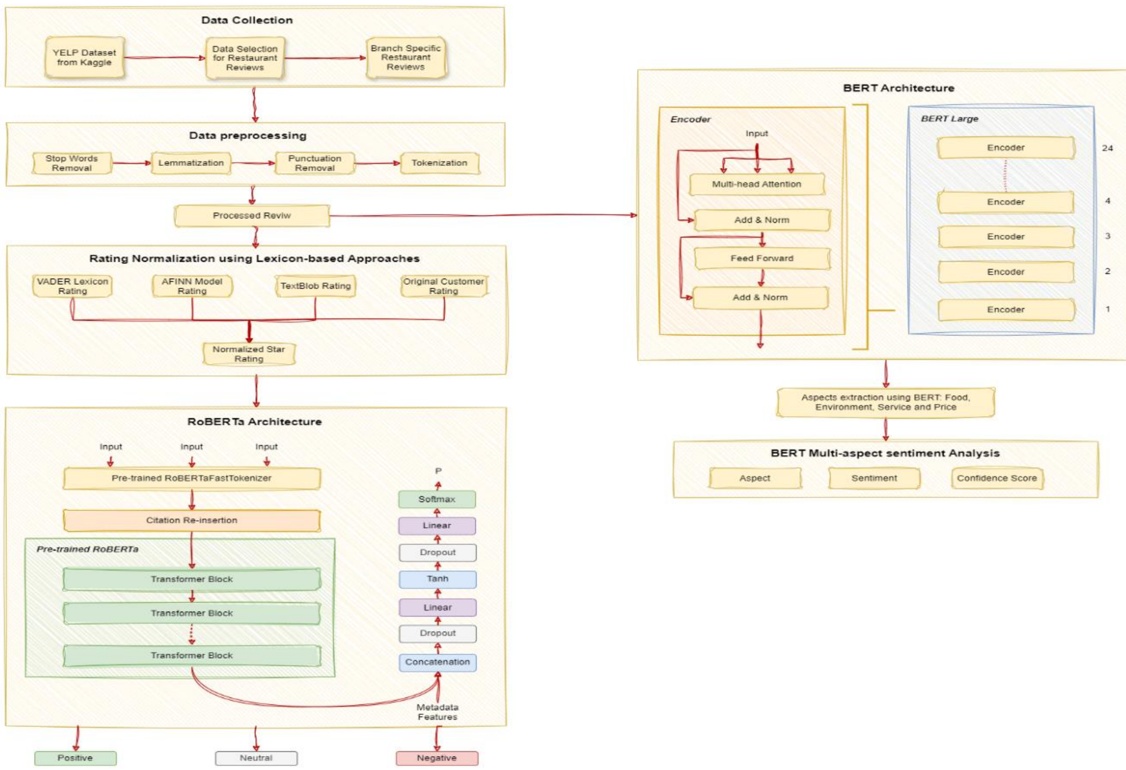

Fig. 1. Hybrid Approach for Hierarchical Spatiotemporal ABSA of Branch specific restaurant.

B. Lexicon-Based Sentiment Analysis

1) VADER Lexicon: VADER, short for Valence Aware Dictionary and sEntiment Reasoner, employs a pre-built lexicon that assigns sentiment scores to individual words. This lexicon is not only focus on the polarity of words but takes into account additional aspects such as their intensity and valence. The comprehensive nature of this lexicon allows VADER to determine the complex emotional embedded in each word, facilitating a more advanced examination of sentiment in textual data [32].

In the VADER methodology, a crucial step involves computing a composite score for each review, serving as an aggregate measure of sentiment. This score considers the collective impact of individual words and their respective sentiment scores, providing an overall evaluation of sentiment within the entire review. A positive score indicates a general positive sentiment, while a negative score suggests a predominant negative sentiment. Scores close to zero imply a more neutral or balanced sentiment, reflecting the complex interaction of emotions expressed in the text.

This approach makes VADER a powerful tool for sentiment analysis, especially in contexts where the detailed and contextual understanding of sentiments is essential. By factoring in intensity and valence alongside polarity, VADER ensures a more refined interpretation of sentiment in textual data, contributing to a deeper comprehension of the emotional subtleties expressed in reviews.

2) AFINN Model: AFINN, a sentiment analysis tool developed by Finn Årup Nielsen, is based on a pre-built list of English words, each assigned a numerical score reflecting its sentiment polarity. With scores ranging from -5 (indicating a highly negative sentiment) to +5 (indicating a highly positive sentiment), AFINN offers a simple yet effective way to measure sentiment in textual data [33]. Unlike more complex models, AFINN's simplicity depends on individual words rather than complex linguistic structures. This approach makes it computationally efficient and particularly suitable for applications where a quick evaluation of sentiment is essential.

In the context of this research, AFINN plays a crucial role in the lexicon-based sentiment analysis of customer reviews. By employing AFINN, we use its existing sentiment scores to assess the emotional tone of words in the reviews. This efficient process aligns with the objective of this research to comprehensively analyze sentiments across a multitude of reviews, contributing valuable insights into client perceptions. The use of AFINN, in aggregation with other sentiment analysis methodologies, improves the depth and accuracy of this sentiment analysis process, to capture the wealth of customer sentiments expressed in the diverse array of reviews from online platforms like Yelp.

3) TextBlob: TextBlob is a Python library that simplifies natural language processing tasks, including sentiment analysis. It employs a pre-trained sentiment analysis model to sort text into different sentiment categories such as positive, negative, or neutral. Similar to VADER and AFINN, TextBlob allows for the quick and efficient analysis of sentiments within textual data.

In the context of this study, TextBlob contributes to a comprehensive sentiment analysis approach by providing an additional layer of analysis. By integrating TextBlob, the capacity to identify subtle sentiments is improved within customer reviews. TextBlob's integral capability to understand and categorize sentiment in a more delicate manner adds depth to this sentiment analysis, ensuring a more refined interpretation of customer sentiments [34]. This multi-method sentiment analysis strategy, combining the strengths of tools like VADER, AFINN, and TextBlob, is essential to extract a unified understanding of the diverse sentiments expressed in customer reviews, thereby contributing to a more thorough examination of the customer experience.

The scores computed by VADER, AFINN, and TextBlob are subsequently combined within a classification model. This model is designed to determine whether the sentiment associated with the input occurrence leans towards positivity or negativity. By unifying the sentiment scores from these three models, a more comprehensive and detailed sentiment classification system is constructed.

C. Advanced Transformer Models for Sentiment Analysis

1) BERT: BERT, or Bidirectional Encoder Representations from Transformers, represents an evolutionary progression in NLP that has significantly developed the understanding and analysis of textual data. Developed by Devlin et al. in their influential work titled "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding" [35], BERT operates on an advanced neural network architecture. What distinguishes BERT is its ability to consider the context of a word within a sentence from both directions, capturing bidirectional context and contextual information. This pre-training on an extensive dataset enables BERT to achieve a profound understanding of language complexities, making it a powerful tool for various NLP tasks.

In this methodology, the capabilities of BERT were exploited to systematically extract essential attributes, including Food, Service, Price, and Environment, from textual reviews. The pre-trained BERT model serves as the backbone, to delve into the complex aspects within reviews and categorize them based on their sentiment. Following the principles outlined in Devlin et al.'s work, BERT-based approach not only extracts specific aspects from reviews but also classifies these aspects into sentiment categories (positive, neutral, and negative). The confidence scores associated with each aspect guide this classification process, resulting in a more detailed and comprehensive representation of sentiments expressed in the analyzed reviews. Overall, BERT's advanced capabilities enhance the sophistication of this sentiment analysis, providing a deeper insight into the multifaceted opinions expressed in textual data.

2) RoBERTa: RoBERTa, another advanced neural network model, improves the bidirectional encoding approach by emphasizing robust pre-training methodologies [36]. Its integration into this hybrid model is designed to augment its capacity to recognize delicate sentiment patterns, particularly in perplex or diverse textual occurrences. By refining the bidirectional encoding approach, RoBERTa contributes to the model's ability to capture complex distinctions in sentiment expression, further enriching the sentiment analysis process. To achieve this objective, the dataset was divided into proportions of 80:10:10, designated for the training, validation, and testing subsets, respectively. This partitioning approach allows for a comprehensive evaluation of the model's performance on unseen data, minimizing the potential for overfitting and underfitting issues.

The integration of both BERT and RoBERTa into this fusion introduces an extra layer of sophistication, empowering the model to achieve a better understanding of sentiment across a diverse range of textual contexts. This approach ensures that the sentiment analysis is not only complete but also adaptable to the complexities present in various forms of textual data essential.

D. Hierarchical Algorithm

In this study, we employed an innovative algorithm for the Hierarchical Spatiotemporal Aspect-based Sentiment Analysis (HISABSA). Specifically designed for unraveling complex layers of sentiment within online reviews, this algorithm is a pioneering approach, leveraging advanced NLP techniques to delve into the spatiotemporal dynamics of sentiment trends and aspect-based categorization. Developed to explore the customer perceptions in the dynamic area of restaurant reviews, HISABSA serves as a cutting-edge tool for extracting valuable insights.

This approach is precisely detailed in Algorithm 1, illustrating a systematic pipeline. Starting with the YELP dataset (D), we applied an iterative algorithm that progresses through hierarchical geographical levels, moving from states to cities and ultimately to individual branches. For each unique branch, reviews (d) within a defined time frame (T) and state (S) were processed. Factors like the time period (T) affect temporal scope, with shorter periods capturing immediate trends and longer ones providing broader perspectives.

The geographical scope defined by the list of states/regions (S) expands analysis breadth but may lead to data sparsity, while the list of cities for each state (C) determines granularity, enhancing detail but increasing computational demands. Branch-specific parameters, like branch size or popularity thresholds, can also impact the reliability of sentiment analysis results locally. Considering these factors can enhance the accuracy of sentiment trends and aspect categorization within each branch.

Following this processing, star ratings were computed using lexicon-based methods, specifically AFINN, VADER, and Textblob. To neutralize potential bias in the original ratings, these derived ratings were normalized by incorporating the original rating. By employing the normalized ratings specific to individual branch reviews, a RoBERTa-based approach is utilized for refined classification. Concurrently, employing a BERT-based methodology, aspects were extracted from the reviews, and associated confidence scores were calculated. The combination of BERT's aspect extraction and lexicon-based rating normalization offers a comprehensive insight into sentiment patterns within restaurant chains.

Algorithm 1: Hierarchical Spatiotemporal Aspect-based Sentiment Analysis (HISABSA)

*1.**Input:** **D** (YELP Dataset), **T** (Time period), **S** (List of states/regions), **C** (List of cities for each state)*
*2:**Output:** Hierarchical sentiment, aspect-based categorization, and rating trend **RT** for branches in each city within each state during time **T**.*
3: D_load ← LOAD(D)
4: T_set ← T
5: D_restaurant ← FILTER(D_load, "restaurant")
6: for s ∈ S do
7: Ds ← FILTER(D_restaurant, s)
8: for c ∈ C do
9: $D_{s,c}$ ← FILTER($D_{s,c}$)
10: for b ∈ B do —▷ Iterate through branches in city c
11: ($D_{s,c,b,T}$)← FILTER($D_{s,c,b,T}$)
12: for d ∈ ($D_{s,c,b,T}$) do
13: dprocessed ← PROCESS(d)
14: R ← RoBERTa_EXTRACT RATINGS(dprocessed)
15: A ← BERT_EXTRACT ASPECTS(dprocessed)
16: SAb ← ANALYZE SENTIMENT($D_{s,c,b,T}$)
end for — ▷ Branch-specific sentiment analysis
17: RT ← CALCULATE REVIEW TREND(R, T)
18: end for
19: end for
20:end for
21: HS_b ← GROUP(SA_b)
22 DISPLAY/STORE HS_b
23:RETURN HS_b, RT

- Input: The algorithm takes the following inputs D (YELP Dataset)
  - ✓ T (Time period): The specific time frame under consideration.
  - ✓ S (List of states/regions): The geographical regions or states relevant to the analysis.
  - ✓ C (List of cities for each state): The cities within each state.
- The Yelp dataset is loaded (D_load), and the subset related to restaurants (D_restaurant) is filtered.
- The algorithm iterates through each state (S) and, within each state, iterates through the cities (C).
- For each city, the algorithm iterates through the branches (B).
- Reviews are filtered based on the specified time period (T) for each branch in each city within each state.
- Each review is processed (dprocessed) as a pre-processing step.
- RoBERTa is employed to extract ratings (R) from the processed review (Line 14).
- BERT is used to extract aspects (A) from the processed review (Line 15).
- The sentiment of each branch (SAb) is analyzed based on the processed reviews (Line 16).
- The rating trend (RT) is calculated based on the extracted ratings and the specified time period (Line 17).
- The results are grouped based on sentiment (HSb) and then either displayed or stored for further analysis (Lines 21-22).
- Output: The algorithm returns the hierarchical sentiment (HSb) and the calculated rating trend (RT) (Line 23).

In essence, this algorithm meticulously examines Yelp reviews, conducting sentiment analysis adapted to individual branches, taking into account distinct time frames and geographic locations. The outcomes are subsequently organized, offering insights into sentiment patterns and the factors shaping customer perspectives across various restaurant branches. This method enables a detailed understanding of customer sentiments, facilitating informed decision-making and strategic improvements fitted to each branch's unique context.

IV. Data Processing Pipeline for Sentiment Analysis

In this data processing pipeline for sentiment analysis, a systematic approach is employed to extract valuable insights from textual data. Employing advanced NLP techniques, the data was preprocessed to ensure its cleanliness and consistency, removing noise and irrelevant information. Through tokenization and embedding, raw text is transformed into a format suitable for analysis models.

Additionally, feature engineering methods are incorporated to capture essential characteristics of the text, improving the accuracy and relevance of this sentiment analysis. This pipeline integrates seamlessly with state-of-the-art sentiment analysis models, facilitating the extraction of detailed sentiments and insights from the data, and to provide decision-makers with actionable intelligence.

A. Data Collection

The cornerstone of this research is the Yelp Dataset Challenge (YDC), a rich collection of business reviews that is readily available on both Kaggle and Yelp's official website. This dataset, sourced from Kaggle, comprises various JSON files including business profiles, user data, reviews, check-ins, and tips, offering a comprehensive perspective on consumer experiences and feedback within the Yelp network. This extensive dataset serves as the primary resource for this investigation to extract various aspects of customer sentiment and behavior in the context of business reviews.

Fig. 2 illustrates the distribution of reviews over the years, highlighting restaurants as the most prevalent category with reviews. Consequently, we have opted to concentrate this study on the restaurant category, given its substantial presence and significance within the dataset.

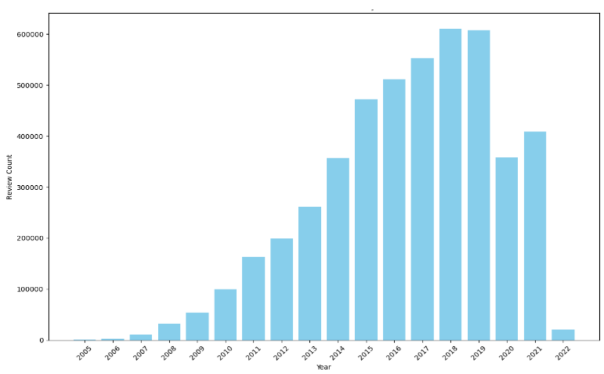

Fig. 2. Year wise restaurant reviews.

B. Data Cleaning and Filtering

In the preliminary phases of this data analysis, significant importance is placed on foundational procedures in data cleaning and filtration to enhance the quality and applicability of the extracted data. This involved a methodical approach to eliminating irrelevant elements such as stop words, numerical figures, and punctuation symbols from the dataset, with the overarching goal of minimizing interference and accentuating substantive content. Through the particular method the dataset underwent a process of refinement that guaranteed subsequent analytical actions, and provided the most pertinent and contextually meaningful textual data [37].

C. Lemmatization

Lemmatization is a linguistic process used in NLP to reduce words to their base or root form, known as a lemma, which helps in standardizing and simplifying text analysis. Unlike stemming, which reduces words to their root form without considering the meaning of the word as illustrated in Fig. 3, lemmatization considers the context and meaning of words resulting in more accurate transformations [38].

The primary objective of lemmatization is to transform inflected or derived words into their base form to facilitate text analysis and improve the accuracy of downstream NLP tasks such as sentiment analysis, information retrieval, and machine translation. For example, lemmatization would convert words like "running" to "run," "better" to "good," and "meeting" to "meet". Lemmatization relies on dictionaries and structural analysis to determine the base form of a word. It considers the part of speech (POS) of each word and applies specific rules to accurately derive the lemma. For instance, verbs are lemmatized to their infinitive forms, nouns to their singular forms, and adjectives to their positive forms.

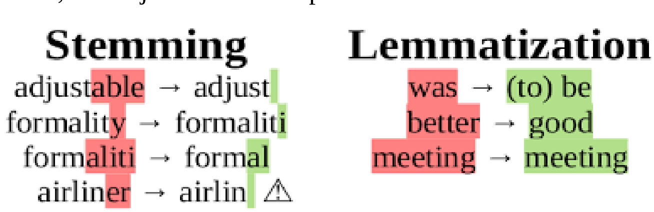

Fig. 3. Stemming vs. Lemmatization.

One advantage of lemmatization over stemming is its ability to produce valid words, as it ensures that the transformed words are present in the dictionary [39]. This helps maintain the semantic integrity of the text and improves the interpretability of the analysis results. In summary, lemmatization is a crucial preprocessing technique that improves the accuracy of text analysis and plays a vital role in various NLP applications by standardizing text data and improving the quality of linguistic analysis.

D. Tokenization

The tokenization of reviews played a pivotal role in increasing the interpretability and analysis of textual data. Tokenization involved segmenting the reviews into meaningful units, such as individual words or phrases, providing a structured foundation for further analysis [40]. Breaking down the text into these discrete units allowed for a more granular understanding of the language used in customer reviews, enabling a detailed exploration of sentiments and opinions.

This process not only facilitated the identification of key terms but also enabled a more accurate examination of linguistic patterns and trends. The use of tokenization as illustrated in Table II. It was instrumental in preparing the textual data for subsequent analyses, contributing to the overall effectiveness of this study by providing a structured and comprehensive source for the exploration of customer experiences across various restaurant branches.

TABLE II. BRANCH SPECIFIC RESTAURANTS PROCESSED REVIEWS

Original Reviews	***Processed Reviews***
We were a bit weary about trying the Shellfish...	bit, weary, trying, Shellfish...
I love trying fresh seafood on piers, wharfs a...	love, trying, fresh, seafood, piers, , wharfs...
Super delish!! No frills! Just great sea food,...	Super, delish, frills, great, sea, food...
For a seafood restaurant at the edge of pear expectations…	seafood, restaurant, edge, pear, expectations...

E. Data Refinement

Exclusively extracted restaurant-related reviews from the dataset, and this improved dataset went through further filtering according to predetermined time frames (T) and geographical states (S), specifically targeting select cities (C) within each state. The goal of this detailed filtering as explained in the HISABSA algorithm, was to ensure that this analysis remained focused and directly pertinent to businesses falling under the restaurant category. This process, illustrated in Fig. 4 increases the precision and relevance to meet research objectives.

However, it is worth mentioning the substantial variability in the quantity of reviews across different restaurants. This diversity underscores the broad spectrum of customer engagements perceived among various restaurants. The primary attention is directed towards the top five restaurant chains: McDonald’s, Subway, Taco Bell, Burger King, and Wendy’s. Fig. 5 supplements this analysis by visually depicting the branch distribution for each of these leading restaurants. This data augments the understanding by providing insights into both the volume of reviews and the geographic dispersion of branches among these prominent establishments.

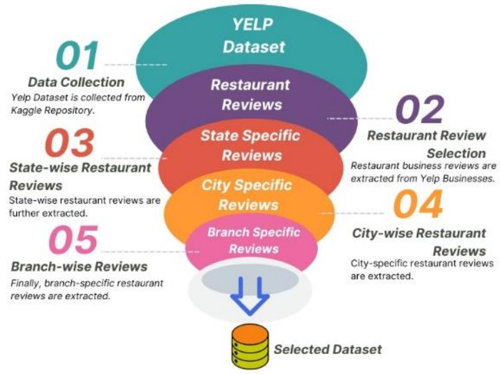

Fig. 4. Branch specific dataset collection.

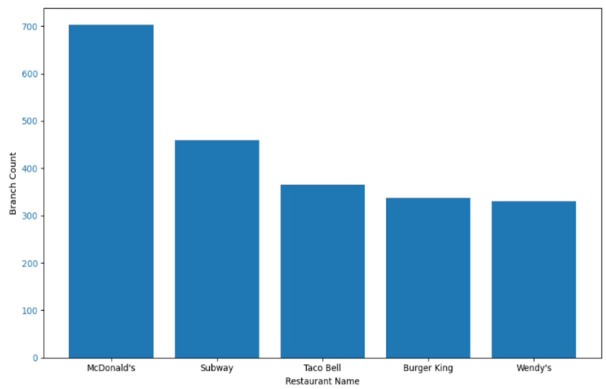

Fig. 5. Branch wise count of top 5 restaurant businesses.

In this study, the analysis was accurately structured to provide a thorough examination of restaurant data, categorizing it systematically by the state and city of each establishment. This methodical approach facilitated a detailed investigation into the individual branches of every restaurant, allowing the collection and analysis of reviews tailored to each location over time. Through Fig. 5 and Fig. 6, the accumulation of customer feedback for every branch of a restaurant chain over its operational lifetime was visually depicted.

Additionally, this study revealed that McDonald's has the highest number of branches compared to other restaurants. This finding underscores McDonald's widespread presence and popularity across various regions. Furthermore, it was found that McDonald's also had the highest number of reviews compared to other restaurants in this analysis. This observation can be attributed to McDonald's status as a globally recognized fast-food chain, attracting a larger customer base and generating more feedback due to its widespread popularity and availability.

Fig. 7 further supports the findings illustrated in Fig. 6 demonstrating that among the top five restaurants, McDonald's stands out with the largest number of reviews, totaling 17,359. This notable volume of reviews underscores McDonald's prominence in customer interactions and feedback, highlighting its significant presence in the restaurant industry. Through this systematic categorization and analytical approach, the accuracy and depth of the findings are enhanced, providing valuable insights into the customer experience across various restaurant locations in diverse regions. This structured analysis enables capturing detailed trends and patterns within the customer feedback landscape, highlighting the dynamics of customer engagement and satisfaction levels across different branches of restaurants.

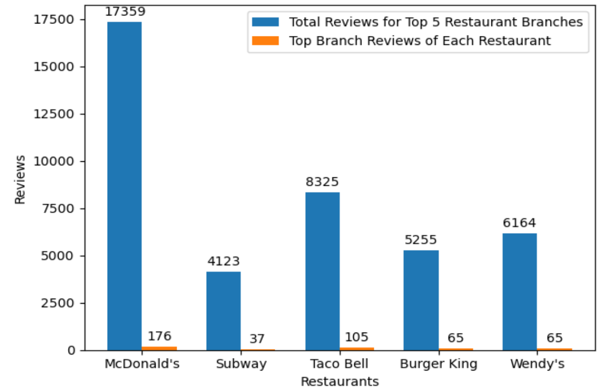

Fig. 6. Reviews count of all restaurant branches and top branch reviews.

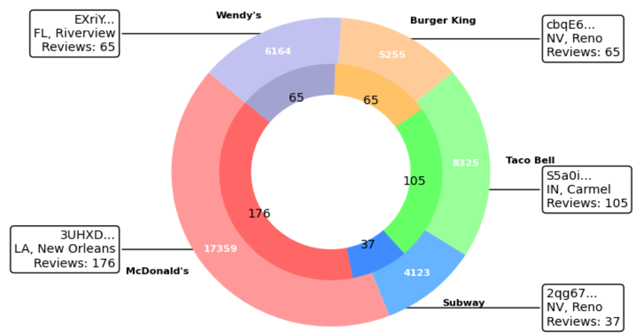

Fig. 7. Reviews count of all restaurant branches and top branch within a particular state and city.

F. *Experimental Setup*

In our research of developing sentiment classification within the area of service industry reviews, we opted for the adoption of a pre-trained RoBERTa model. This decision was driven by the model's robust linguistic comprehension, enabling it to interpret a diverse dataset consisting of reviews from various McDonald's branches. The computational backbone of this framework was the NVIDIA A100 GPU, boasting an ample 80 GB of memory, which played a pivotal role in accelerating both training and experimentation phases.

This hardware choice proved crucial in managing the scale of the data and the computational demands essential in fine-tuning RoBERTa for this specific task requirements. Additionally, a BERT-based approach is implemented for aspect-based sentiment analysis, focusing on sentiment trends. The computational infrastructure supporting the BERT-based methodology included high-performance computing resources, notably featuring multi-core processors and sufficient RAM capacity to handle the computational demands of processing large volumes of textual data.

Moreover, cloud-based platforms equipped with robust GPU acceleration capabilities were leveraged, enhancing the efficiency and speed of the aspect extraction and sentiment analysis tasks. This approach successfully condensed the overall general aspects of a review into these specific categories, along with the assignment of confidence scores.

V. Unveiling Sentiment Analysis: From Lexicons to Transformers

A. Refining Sentiment Analysis: VADER, AFINN, and TextBlob Integration

This study employed the VADER sentiment analysis tool to evaluate and quantify the sentiments expressed in customer reviews. This process uses the VADER analyzer that starts with the calculation of sentiment percentages for each review, breaking down the sentiments into positive, negative, and neutral components. These values were then utilized to derive an average sentiment score, referred to as the 'VADER rating'.

Additionally, statistical summary metrics such as maximum, minimum, and average values are computed to offer a comprehensive understanding of the sentiment distribution in the dataset. The categorization of sentiment scores further improves the interpretability of the results, providing a more detailed perspective on customer sentiments. The systematic integration of the VADER tool into this analysis framework contributes to a robust and quantifiable assessment of sentiment trends in the context of restaurant reviews [41].

Moreover, AFINN sentiment analysis tool is employed to measure the overall sentiment in processed reviews, with a particular focus on positive scores that were assumed correlated with positive sentiments. The intentional classification of these positive scores into five specific groups was carefully designed to include a detailed range similar to a star rating system, covering from 1 star to 5 stars. This categorization provides a more granular understanding of sentiment distribution, allowing the differentiation between various levels of positivity in the dataset.

AFINN's attributes simplicity, speed, linguistic approach, and interpretability, make it able for typical sentiment analysis tasks. Its unified implementation and efficiency in handling large datasets facilitate quick assessments of sentiment polarity. Ultimately, the practical application of AFINN in this sentiment analysis is guided by its efficiency in swiftly providing insights into sentiment polarity, aligning with a simplified yet meaningful star rating framework.

The TextBlob library is utilized to perform sentiment analysis on the processed reviews. The code extracts polarity and subjectivity scores using TextBlob's sentiment analysis capabilities, providing insights into the sentiment's positivity (polarity) and its level of objectivity [42]. This score is a numerical representation of the sentiment, computed as the division of polarity by subjectivity aspects as shown in Eq. (1), incorporating a small constant in the denominator to avoid division by zero. The subsequent categorization of the scores into five classes, based on predefined entries, contributes to a detailed understanding of sentiment distribution within the dataset. The resulting 'TextBlob rating' represents these categorized sentiment scores.

$$Score = \frac{polarity}{subjectivity+\epsilon} \quad (1)$$

The study further enriches its sentiment analysis by integrating TextBlob's results with other sentiment scores, including those derived from AFINN and VADER, as well as the original star ratings provided by the customers. This integrated approach aims to offer a comprehensive and multi-faceted evaluation of sentiment.

B. Exploring Sentiment Analysis with RoBERTa and BERT Transformers

A systematic framework is established for this approach, incorporating revolutionary DL models, namely RoBERTa and BERT, in an innovative manner. Both models operate on transformer-based architectures, representing NLP advancements remaining to their deep learning structures. Their bidirectional training proves particularly effective for sentiment analysis, capturing sensitive implications embedded in word order and sentence structure that significantly influence sentiment. Adjusted to this specific dataset, these models excel in sentiment analysis, serving as foundational tools that provide scalable solutions adaptable to a diverse range of languages and domains.

In this research, BERT was initially employed as a powerful transformer-based model, to process and analyze textual data from customer reviews. BERT played a crucial role in this methodology by extracting important aspects from the text, such as food quality, service standards, and pricing perceptions. Its sophisticated architecture allowed the categorization of these aspects and effectively captured and classified diverse elements within the reviews, contributing to a comprehensive understanding of customer feedback and preferences.

Following the initial analysis with BERT, RoBERTa was integrated into this methodology for sentiment classification. RoBERTa, another transformer-based model, that excels in understanding and interpreting the semantic distinctions present in text data. In this research, RoBERTa was assigned to the pivotal role of sentiment classification within the reviews. By exploiting its robust linguistic comprehension and sophisticated algorithms, RoBERTa effectively categorized the sentiments expressed by customers as positive, negative, or neutral. The integration of RoBERTa enhanced the accuracy and depth of this sentiment analysis, providing valuable insights into the overall sentiment trends across different restaurant branches and customer interactions [43].

The dataset is balanced by down sampling the negative class and then shuffled to create a balanced dataset. The RoBERTa model is utilized to create a neural network with bidirectional Long Short-Term Memory (LSTM) layers, which are optimized on the balanced dataset. Moreover, the model is then compiled using the AdamWeightDecay optimizer and categorical cross entropy loss function. The model's architecture includes bidirectional LSTM layers followed by dense layers with dropout and batch normalization. The RoBERTa embeddings are used as input to these layers, then the model is trained on the prepared training and validation data, then its performance is evaluated using the Categorical Accuracy metric.

Finally, a tokenizer specific to RoBERTa is employed to preprocess text data for predictions. The 'predict' function takes a text input, preprocesses it, and utilizes the trained model to predict the sentiment, returning the corresponding index of the predicted sentiment category (Positive, Negative, or Neutral). The study aims to employ the capabilities of transformer-based models like RoBERTa to achieve robust sentiment analysis results on the provided dataset.

VI. RESULTS

A. Evaluation of Lexicon based Approaches

In this section, the sentiment analysis is augmented by integrating outcomes from VADER, AFINN, and TextBlob, in addition to the original star ratings provided by customers. This comprehensive approach ensures a detailed understanding of customer feedback, transcending the limitations of numerical ratings and offering a more comprehensive representation of their sentiments. Through a meticulously structured analytical process outlined in Table III, each of the three methodologies contributes valuable insights to the overarching sentiment analysis.

TABLE III. NORMALIZED EXTRACTED RATING USING LEXICON BASED APPROACHES

Original Reviews	*Original Rating*	*VADER Rating*	*AFINN Rating*	*TextBlob Rating*	*Normalized Extracted Rating*
We were a bit weary about trying the Shellfish...	4	2	5	4	3.6
I love trying fresh seafood on piers, wharfs a...	4	2	5	4	3.6
Super delish!! No frills! Just great sea food,...	4	4	2	4	3.3
For a seafood restaurant at the edge of pear expectations…	3	2	4	3	3.0

Acknowledging that numerical ratings may lack granularity in conveying specific preferences or concerns, these methodologies were employed to extract normalized ratings from the textual content of reviews. This method aims to provide a more pertinent and insightful understanding of what aspects customers appreciate. The strength of these results lies in their ability to capture detailed sentiments expressed in textual reviews, thereby enriching the analysis and offering deeper insights into customer preferences and concerns.

B. Evaluation of Transformer Models

The results of the normalized extracted ratings were leveraged as a training dataset for the BERT and RoBERTa models. The dataset was partitioned into an 80% training set, a 10% validation set, and a 10% test set. This strategic division allowed training the model on a substantial portion of the data, validate the performance on a separate set to adjusted parameters, and ultimately assess the generalization on an independent test set. The normalized ratings, derived from the lexicon-based methodologies and reflecting sentiments from customer reviews, served as valuable inputs for training our model. This approach aimed to expand the models' understanding of sentiment distinctions and improve the accuracy in predicting sentiment labels across various customer feedback scenarios.

Conducting a Hierarchical Spatiotemporal Aspect-Based Sentiment Analysis (HISABSA), the city of Indianapolis was selected since it hosts a substantial number of reviews with relevant aspects. The dataset includes 176 customer reviews for a McDonald's branch in Indianapolis, Indiana. The primary aspects of Food, Service, Environment, and Price were carefully extracted for analysis. The findings are visually represented in Fig. 8 and Fig. 9 showcasing the temporal trends for each aspect and the overall rating trend for complete reviews over time, respectively. Upon visually interpreting these two figures, a discernible shift in sentiment across various aspects of the restaurant branch becomes evident.

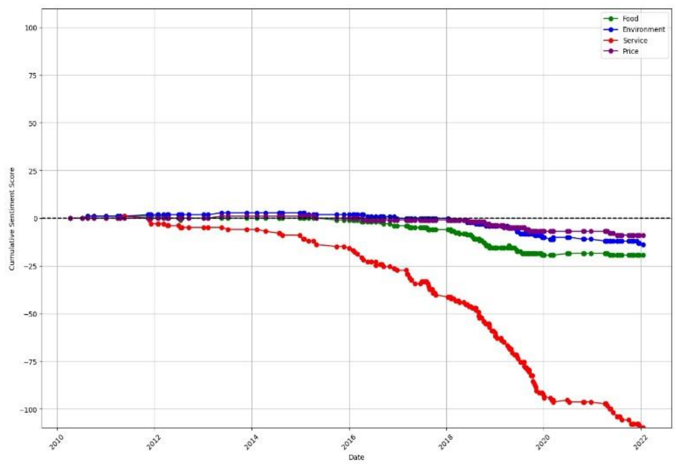

Fig. 8. Cumulative sentiment trends of restaurant branch aspects in Indianapolis, Indiana over time.

Fig. 8 illustrates the cumulative sentiment trends related to Food, Environment, Service, and Price, revealing a substantial decline in sentiment scores, particularly noticeable from 2018 to 2022. This trend signals a change in customer perceptions regarding different aspects of the restaurant branch during this period. The decrease in sentiment scores implies possible issues or alterations in customer experiences, prompting the need for deeper exploration into the factors driving these changes. Moreover, analyzing the specific time frames corresponding to sentiment score decreases can provide valuable insights into areas for potential improvement or refinement within the restaurant's operations and services.

Upon thorough examination, it's clear that the 'Service' aspect undergoes the most significant decline in sentiment, indicating a intensified sensitivity among customers towards the perceived quality of service during this period. In contrast, 'Food' and 'Environment' show a more gradual decline, while 'Price' demonstrates relatively lower volatility but still trends negatively. Addressing these findings necessitates a comprehensive strategy that encompasses all aspects of the restaurant's offerings. Potential interventions may involve reviewing pricing strategies, re-evaluating menu items and food quality, improving the dining ambiance, and investing in customer service training.

Fig. 9 serves as a supplementary illustration to the initial findings, presenting the trajectory of overall star ratings over time. Similarly, a noteworthy concentration of lower star ratings is evident, specifically occurring the latter part of the period between 2018 and 2022. This alignment implies a consistent decline not only in specific aspects of the customer experience but also in overall satisfaction. The temporal correlation implies that the restaurant faced challenges during this timeframe, influencing various facets of its operations. The simultaneous decline across all four aspects indicates that the negative sentiment was not confined to a singular area but rather filtered different elements of the dining experience.

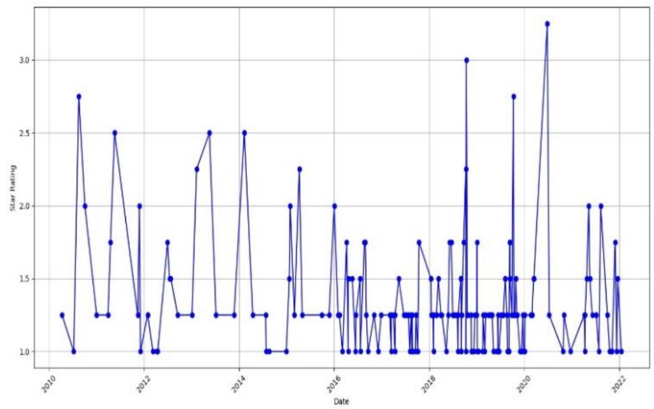

Fig. 9. Star rating trend of restaurant branch in state in, city, Indianapolis over time

The research aims to comprehensively understand both internal operations and external influences shaping customer perceptions. Through this complete interpretation, the restaurant can devise an effective plan for recovery and enrichment. By delving deeper into customer sentiments and preferences, businesses can orient their approaches to improve satisfaction, foster brand loyalty, and ultimately sustain profitability. Exploiting the insights gathered from this research, companies can navigate competitive markets with informed decisions that drive growth and success.

C. Evaluation Metrics

The dataset underwent thorough preprocessing to align with the input expectations of the RoBERTa architecture, including the normalization of star ratings to serve as labels in supervised learning. The training procedure was fine-tuned to optimize the model's performance across three sentiment classes, representing a range of customer opinions. Rigorous validation ensured the robustness and applicability of the model, with these findings illuminated by a confusion matrix showcasing the model's differentiation among sentiment classes.

The evaluation of the refined RoBERTa model demonstrated praiseworthy precision, recall, and F1-score metrics, averaging approximately 0.92, 0.93, and 0.90, respectively, across sentiment classes, resulting in an overall accuracy of 92%. These outcomes represent a significant enhancement over existing state-of-the-art models in sentiment analysis within the restaurant review context. The confusion matrix in Fig. 10 further substantiates the model's efficacy, illustrating a high degree of accuracy in class predictions with minimal misclassification.

Table IV illustrates that the model reveals its ability to accurately identify positive, negative, and neutral sentiments expressed in customer reviews. High recall values, notably reaching 0.98 for negative sentiment, signify the model's proficiency in retrieving relevant instances of each sentiment class from the dataset. The F1-Score values, ranging from 0.90 to 0.95, reflect a balanced trade-off between precision and recall, indicating the model's robustness in handling sentiment classification tasks. These findings affirm the HISABSA approach as a robust and accurate framework for sentiment analysis, offering valuable insights into customer perceptions and feedback within the restaurant industry.

The HISABSA study stands out among other sentiment analysis approaches, as demonstrated by the results summarized in Table V. While previous studies have achieved varying levels of performance across different techniques on the YELP dataset, our approach consistently outperformed these studies in terms of accuracy and F1-score. For instance, in one study [31] using the YELP dataset, techniques such as Logistic Regression, Naive Bayes, and Support Vector Clustering returned accuracies ranging from 73.22% to 79.12%. Furthermore, in another experiment utilizing the YELP dataset, various techniques including Multinomial Naive Bayes and Support Vector Machine achieved F1-scores ranging from 0.701 to 0.757. In contrast, the HISABSA approach utilizing RoBERTA achieved an impressive accuracy of 92%, surpassing the performance of other methods and showcasing its competitive performance against traditional methods.

These results highlight the effectiveness of this study's approach in sentiment analysis, particularly in the context of restaurant reviews. By leveraging advanced techniques such as RoBERTA and fitting this methodology to the specific characteristics of the dataset, we were able to achieve superior performance in accurately classifying sentiments expressed in customer reviews. This underscores the robustness and reliability of the HISABSA framework in extracting valuable insights from textual data and contributing to a deeper understanding of customer sentiments in the restaurant industry.

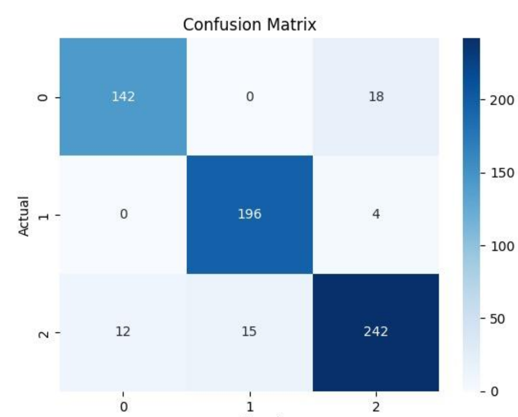

Fig. 10. Confusion matrix of the RoBERTa classification model.

TABLE IV. RESULTS OF THE ROBERTA BASED APPROACH TO CLASSIFY RESTAURANT REVIEWS

Sentiment Class	*Precision*	*Recall*	*F1-Score*	*Accuracy*
0 (Positive)	0.92	0.89	0.90	0.92
1 (Negative)	0.93	0.98	0.95	0.92
2 (Neutral)	0.92	0.90	0.91	0.92

TABLE V. A COMPARISON OF SENTIMENT ANALYSIS APPROACHES FOR ONLINE REVIEWS

Ref	*Dataset*	*Technique*	*Performance (f1-score)*
[27]	YELP	LR NB CNN BERT ALBERT	83.72% 73.35% 87.85% 89.47% 89.21%
[28]	YELP	SVM RF Multinomial NB KNN	0.76% 0.78% 0.77% 0.61%
[29]	YELP	Multinomial Naive Bayes SVM Logistic Regression Gradient Boosting (XGBoost) BERT LSTM	F1-score: 0.731 0.757 0.757 0.750 0.723 0.701
[30]	YELP	SVM	Accuracy: 88%
[31]	YELP	Logistic Regression Naive Bayes Bernoulli Naive Bayes Multinomial Naive Bayes Linear SVC (Support Vector Clustering) Proposed	Accuracy: 78.88% 79.12% 73.22% 78.92% 75.32% 78.44%
HISABSA (this study's Approach)	**YELP**	**Hybryd Sentiment Analysis Architecture (RoBERTA)**	**Accuracy: 92%**

VII. DISCUSSION

The strength of this study lies in its comprehensive approach towards understanding customer sentiments within the restaurant industry, addressing the challenges inherent in sentiment analysis and offering valuable insights for managerial decision-making. By introducing the Hierarchical Spatiotemporal Aspect-Based Sentiment Analysis (HISABSA) methodology, this research bridges the gap between traditional sentiment analysis methods and the complex dynamics of customer feedback in the restaurant sector. Considering both textual content and numerical ratings, our model achieves a sophisticated understanding of customer feedback, leading to more detailed insights and more accurate sentiment analysis results.

Based on the results, the proposed model has been fine-tuned to optimize its parameters, resulting in robust performance that outperforms existing studies. One of the main strengths of HISABSA model is its integration of diverse sentiment analysis methodologies, including lexicon-based approaches alongside advanced transformer models. This hybrid model enables a thorough examination of customer sentiments, capturing subtle distinctions and providing a more accurate interpretation of their experiences.

Moreover, the model's robust performance, as demonstrated by the evaluation metrics, underscores its effectiveness in accurately analyzing and categorizing customer feedback. The high precision, recall, and F1-score metrics, along with the remarkable 92% accuracy in sentiment classification, highlight the model's capability to differentiate among sentiment classes and provide reliable insights for decision-making. This heightened level of accuracy can be attributed to the hybrid model that combines various techniques, leveraging the strengths of each approach to effectively capture diverse aspects of the input data.

The findings of this study highlight a distinct need for improvements across various facets within the restaurants. Systematically addressing the decline in customer sentiment across the aspects and comprehending the root causes will enable the restaurant to elevate customer satisfaction and restore its reputation moving forward. It is imperative for the branch manager to delve deeper into the sentiment of each aspect, correlating them with specific timeframes and external events.

In meeting the challenges faced by the restaurant industry, this study offers actionable insights for improving customer satisfaction and enhancing business performance. By delving into specific aspects such as food quality, service standards, and pricing perceptions, the model provides a complete view of customer sentiments, enabling restaurants to identify areas for improvement and tailor their strategies accordingly. Therefore, it can be concluded that the combination of these fundamental models within a classification model has excelled in facilitating performance enhancement by capturing distinct features of the input data based on their respective operational modes.

VIII. CONCLUSION

In summary, this research aims to uncover the intricacies of customer sentiments within the restaurant industry by employing a comprehensive methodology. It combines lexicon-based approaches, utilizing VADER, AFINN, and TextBlob, with innovative transformer models like BERT and RoBERTa, to gain a thorough understanding of customer feedback. Recognizing the inherent limitations of numerical star ratings provided by customers, leveraging lexicon-based methodologies extracts detailed and standardized ratings from customer reviews. This approach fosters a more refined understanding of sentiments, serving as a crucial training dataset for the transformer models. This integration aims to empower this study's model to capture the subtle complexities of customer sentiments, providing a more accurate interpretation of their experiences.

The primary objective of this study was to develop an advanced hybrid sentiment analysis model capable of accurately analyzing and categorizing customer feedback from restaurant reviews. Conducting an in-depth examination of Hierarchical Spatiotemporal Aspect-Based Sentiment Analysis (HISABSA), this research meticulously analyzes 176 customer reviews from a specific McDonald's branch in Indianapolis, Indiana. This extensive analysis delves into primary aspects such as Food, Service, Environment, and Price, providing a comprehensive investigation. Moreover, the optimized RoBERTa model, trained on the dataset derived from lexicon-based methodologies, achieves creditable precision, recall, and F1-score metrics, outperforming existing state-of-the-art models. The confusion matrix underscores the model's efficacy, highlighting its nuanced differentiation among sentiment classes. The experimental results demonstrate the successful achievement of this objective, with HISABSA the model showcasing a outstanding results in classifying sentiments across positive, negative, and neutral categories.

Visual representations reveal evident changes in sentiment over time, particularly highlighting notable declines in service. The simultaneous decline across all aspects between 2018 and 2022 suggests challenges faced by the restaurant, highlighting the need for comprehensive improvements across various operational facets. To contextualize these results, it is crucial to consider the challenges inherent in sentiment analysis within the restaurant industry. Restaurants face the daunting task of deciphering and understanding the diverse sentiments expressed by customers across various dimensions such as food quality, service standards, and pricing perceptions. The hybrid model developed in this study addresses these challenges by integrating both lexicon-based and context-based approaches, thereby providing a comprehensive understanding of customer sentiments. These findings not only contribute to the advancement of ABSA but also hold significant implications for business intelligence, offering decision-makers clear insights into customer needs and providing a robust framework for decision-making.

In conclusion, the HISABSA findings underscore the importance of systematically addressing specific aspects to develop overall customer satisfaction and restore a restaurant's reputation. By delving into the sentiment of each aspect, correlating them with specific timeframes and external events, restaurant managers can make informed decisions for a comprehensive development of customer satisfaction. This study contributes a unified framework for sentiment analysis in the restaurant industry, offering actionable insights for managerial decision-making, continuous improvement, and sustained success.

In future research, optimizing model parameters, expanding analysis to include multiple languages, and incorporating dynamic sentiment analysis techniques offer opportunities to enhance the hybrid sentiment analysis model's performance and applicability. Integrating user-generated content from diverse sources and incorporating domain-specific knowledge further refine sentiment analysis methodologies for the restaurant industry.

REFERENCES

[1] Zhang, L. and B. Liu, Aspect and entity extraction for opinion mining, in Data mining and knowledge discovery for big data: Methodologies, challenge and opportunities. 2014, Springer. p. 1-40.

[2] Chepukaka, Z.K. and F.K. Kirugi, Service Quality and Customer Satisfaction at Kenya National Archives and Documentation Service, Nairobi County: Servqual Model Revisited. Int. J. Cust. Relat, 2019. **7**(1).

[3] *Tripadvisor*. 2023 [cited 2023 October]; Available from: https://www.tripadvisor.com/.

[4] *Yelp Dataset*. 2023 [cited 2023 October]; Available from: https://www.yelp.com/dataset.

[5] Chauhan, G.S., et al., A two-step hybrid unsupervised model with attention mechanism for aspect extraction. Expert systems with Applications, 2020. **161**: p. 113673.

[6] García-Pablos, A., M. Cuadros, and G. Rigau, W2VLDA: almost unsupervised system for aspect based sentiment analysis. Expert Systems with Applications, 2018. 91: p. 127-137.

[7] Alharbi, M., J. Yin, and H. Wang. Surveying the Landscape: Compound Methods for Aspect-Based Sentiment Analysis. in Databases Theory and Applications. 2024. Cham: Springer Nature Switzerland.

[8] Kaur, G. and A. Sharma, A deep learning-based model using hybrid feature extraction approach for consumer sentiment analysis. Journal of Big Data, 2023. 10(1): p. 5.

[9] Ara, J., et al. Understanding customer sentiment: Lexical analysis of restaurant reviews. in 2020 IEEE Region 10 Symposium (TENSYMP). 2020. Dhaka, Bangladesh: IEEE.

[10] Banjar, A., et al., Aspect-Based Sentiment Analysis for Polarity Estimation of Customer Reviews on Twitter. Computers, Materials & Continua, 2021. 67(2).

[11] Jain, P.K., R. Pamula, and G. Srivastava, A systematic literature review on machine learning applications for consumer sentiment analysis using online reviews. Computer science review, 2021. 41: p. 100413.

[12] Rani, S. and P. Kumar, Deep Learning Based Sentiment Analysis Using Convolution Neural Network. Arabian Journal for Science and Engineering, 2019. 44(4): p. 3305-3314.

[13] Lan You, et al. ASK-RoBERTa: A pretraining model for aspect-based sentiment classification via sentiment knowledge mining. Knowledge-Based Systems. Volume 253, 2022 October 11, 109511.

[14] Aubaid, A.M. and A. Mishra, A rule-based approach to embedding techniques for text document classification. Applied Sciences, 2020. 10(11): p. 4009.

[15] Potisuk, S. Typed dependency relations for syntactic analysis of Thai sentences. in Proceedings of the 24th Pacific Asia Conference on Language, Information and Computation. 2010. Tohoku University, Sendai, Japan.

[16] Ray, P. and A. Chakrabarti, A mixed approach of deep learning method and rule-based method to improve aspect level sentiment analysis. Applied Computing and Informatics, 2022. 18(1/2): p. 163-178.

[17] Asghar, M.Z., et al., Lexicon-enhanced sentiment analysis framework using rule-based classification scheme. PloS one, 2017. 12(2): p. e0171649.

[18] Hutto, C. and E. Gilbert. Vader: A parsimonious rule-based model for sentiment analysis of social media text. in Proceedings of the international AAAI conference on web and social media. 2014. Ann Arbor, Michigan USA.

[19] Birjali, M., M. Kasri, and A. Beni-Hssane, A comprehensive survey on sentiment analysis: Approaches, challenges and trends. Knowledge-Based Systems, 2021. 226: p. 107134.

[20] Hegde, S., S. Satyappanavar, and S. Setty. Restaurant setup business analysis using yelp dataset. in 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI). 2017. Udupi, Karnataka, India: IEEE.

[21] Mubarok, M.S., A. Adiwijaya, and M.D. Aldhi. Aspect-based sentiment analysis to review products using Naïve Bayes. in AIP conference proceedings. 2017. Malaysia: AIP Publishing.

[22] Hammi, S., S.M. Hammami, and L.H. Belguith, Advancing aspect-based sentiment analysis with a novel architecture combining deep learning models CNN and bi-RNN with the machine learning model SVM. Social Network Analysis and Mining, 2023. 13(1): p. 117.

[23] Wang, J., et al. Aspect sentiment classification with both word-level and clause-level attention networks. in IJCAI. 2018.

[24] Wankhade, M., C.S.R. Annavarapu, and A. Abraham, MAPA BiLSTM-BERT: multi-aspects position aware attention for aspect level sentiment analysis. The Journal of Supercomputing, 2023. 79(10): p. 11452-11477.

[25] Khan, M.U., et al., A novel category detection of social media reviews in the restaurant industry. Multimedia Systems, 2020: p. 1-14.

[26] Verma, S., A. Kumar, and A. Sharan, IAN-BERT: Combining Post-trained BERT with Interactive Attention Network for Aspect-Based Sentiment Analysis. SN Computer Science, 2023. 4(6): p. 756.

[27] Alamoudi, E.S. and N.S. Alghamdi, Sentiment classification and aspect-based sentiment analysis on yelp reviews using deep learning and word embeddings. Journal of Decision Systems, 2021. 30(2-3): p. 259-281.

[28] Gupta, K., N. Jiwani, and N. Afreen, A Combined Approach of Sentimental Analysis Using Machine Learning Techniques. Revue d'Intelligence Artificielle, 2023. 37(1).

[29] Liu, S., Sentiment analysis of yelp reviews: a comparison of techniques and models. arXiv preprint arXiv:2004.13851, 2020.

[30] Yu, B., et al., Identifying restaurant features via sentiment analysis on yelp reviews. arXiv preprint arXiv:1709.08698, 2017.

[31] Hemalatha, S. and R. Ramathmika. Sentiment analysis of yelp reviews by machine learning. in 2019 International Conference on Intelligent Computing and Control Systems (ICCS). 2019. Madurai, India: IEEE.

[32] Isnan, M., Elwirehardja, G. N., & Pardamean, B. (2023). Sentiment Analysis for TikTok Review Using VADER Sentiment and SVM Model. Procedia Computer Science, 227, 168-175.

[33] Zezawar, T.K., Aung, N.M., Sentiment analysis of students' comment using lexicon based approach. IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS), 2017.

[34] Manuaba, I. B. K. (2023). A Sentiment Analysis Model for the COVID-19 Vaccine in Indonesia Using Twitter API v2, TextBlob, and Googletrans. Procedia Computer Science, 227, 1101-1110.

[35] Devlin, J., et al., Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805, 2018.

[36] Liu, Y., et al., Roberta: A robustly optimized bert pretraining approach. arXiv preprint arXiv:1907.11692, 2019.

[37] Nielsen, F.Å., A new ANEW: Evaluation of a word list for sentiment analysis in microblogs. arXiv preprint arXiv:1103.2903, 2011.

[38] Nuţu, Maria., Deep Learning Approach for Automatic Romanian Lemmatization, Procedia Computer Science, Volume 192, 49-58, 2021.

[39] Padmanabhan, Arvind. (n.d.). "Arvind Padmanabhan." Retrieved October 11, 2019, from https://devopedia.org/lemmatization.

[40] Souza, F.C., Nogueira, R.F., Lotufo, R.A., BERT models for Brazilian Portuguese: Pretraining, evaluation and tokenization analysis, Applied Soft Computing, Volume 149, Part A, December 2023, 110901.

[41] Borg, A., Boldt, M., Using VADER sentiment and SVM for predicting customer response sentiment, Expert Systems with Applications, Volume 162, 30 December 2020, 113746.

[42] Rosenberg, E., Tarazona, C., Mallor, F., Eivazi, H., Pastor-Escuredo, D., Fuso-Nerini, F., Vinuesa, R., Sentiment analysis on Twitter data towards climate action, Results in Engineering, Volume 19, September 2023, 101287.

[43] Malik, M.S.I., Nazarova, A., Jamjoom, M.M., Ignatov, D.I., Multilingual hope speech detection: A Robust framework using transfer learning of fine-tuning RoBERTa model. Journal of King Saud University - Computer and Information Sciences, Volume 35, Issue 8, 101736, 2023.