NEWCASTLE BUSINESS SCHOOL

Newcastle Business School, The University of Newcastle, NSW, Australia

Course code: BUSA 3001
Course name: AI in Business

Assignment 2 Report

BREWING INSIGNTS: AI-POWERED ANALYSIS OF STARBUCKS CUSTOMER FEEDBACK BY LDA ALGORITHM

Cover PAGE

Name and student number	Section Completed	Signed	Dated
Tan Thuy Linh Dao – C3425175	**Executive Summary** **Introduction** **Background of NLP Features** **Proposed NLP AI Solution Architecture** **Data Presentation** **Discussion and Conclusion** **Reference**	**Linh**	**30th-May**

Word count: 3300 words

TABLE OF CONTENTS

TABLE OF FIGURES

EXECUTIVE SUMMARY

This report demonstrates how NLP-powered topic modeling (LDA) transforms unstructured Starbucks customer reviews into actionable business insights. Key findings highlight dominant customer concerns—service quality, wait times, and staff management—while temporal analysis tracks evolving priorities. Visualisations (word clouds, n-grams, and trend graphs) enable managers to quickly discover pain points and emerging concerns. The solution improves decision-making by automating feedback analysis at scale; however, ethical considerations around data privacy and bias must be addressed. Future integration with real-time dashboards and sentiment analysis could further optimise customer experience strategies. This AI-driven approach enables businesses to proactively align operations with customer expectations.

1. INTRODUCTION

In today's digital economy, businesses are inuncated with unstructured customer feedback from reviews, social media, and surveys (Liu et al., 2020). While rich in insights, this textual data overwhelms traditional analytical methods designed for structured, numerical information (Mishra et al, 2024). The resulting analysis gap poses significant business risks, including missed opportunities and delayed responses in an era when customers expect real-time involvement.

Natural Language Processing (NLP) emerges as a critical solution to this challenge, enabling automated interpretation of human language at scale (Sintoris & Vergidis, 2017). This technology uses techniques like sentiment analysis and topic modeling to transform raw text into actionable business intelligence (Mishra et al, 2024). Latent Dirichlet Allocation (LDA) is a sophisticated unsupervised machine learning algorithm that identifies hidden themes in customer feedback without predefined categories.

This report analyses how NLP-powered automation enhances managerial decision-making using Starbucks customer reviews as a case study. We demonstrate how LDA can evaluate reviews to uncover recurring concerns about service, product quality, or pricing – insights that would be impractical to extract manually. The analysis generates intuitive visualisations including word clouds and topic trends, allowing managers to identify operational improvements quickly.

Beyond operational benefits, the report discusses crucial ethical considerations when deploying AI for customer data analysis. As organisations increasingly employ these technologies, ensuring transparency and fairness in automated decision processes becomes critical to maintaining customer trust.

2. BACKGROUND OF NLP FEATURES

2.1. The Rise of Text Data in Business Context – Starbucks Case Study

The digital revolution has fundamentally transformed how businesses interact with customer feedback. Whereas companies once relied on structured surveys and sales data, they now face a deluge of unstructured text from social media, review sites, and apps (Ittoo et al., 2016). Starbucks exhibits this challenge perfectly. As one of the world's most well-known coffee chains, it receives a constant stream of customer feedback via its mobile app, website, and third-party review sites. Each day, thousands of new comments are made regarding drinks, service quality, and overall customer experience.

This explosion of textual data presents both opportunity and analytical challenges. While these customer voices contain significant insights, traditional manual analysis approaches simply cannot keep pace. Human analysts struggle to process the volume, frequently missing minor patterns and emerging trends buried in the noise. The consequences are severe: delayed replies to customer concerns, missed opportunities for improvement, and an inability to identify issues before they escalate.

2.2. What is NLP and why is it important?

Natural Language Processing (NLP) represents a critical intersection between computer science and linguistics (Flayeh et al., 2022), enabling computers to understand, interpret, and generate human language. Emerging from artificial intelligence research within the broader field of

informatics, NLP systems analyze both written and spoken language to extract meaningful information and actionable insights from unstructured text data (Sintoris & Vergidis, 2017).
NLP's business applications have become increasingly vital in today's data-driven economy. Organisations now leverage NLP-powered text analytics to process enormous volumes of customer feedback, transforming free-form comments into structured data, employ machine learning algorithms to identify hidden patterns, and identify emerging trends within customer opinions (Flayeh et al., 2022). Therefore, companies can quickly detect customer sentiment shifts, prioritise improvement areas, and generate visual reports for management decision-making.
Latent Dirichlet Allocation (LDA) is a widely used NLP approach for customer feedback analysis. It operates as a generative probabilistic model that represents documents as mixtures of latent topics (Jelodar et al., 2019), automatically discovers the underlying themes and concerns without requiring pre-defined categories. These topics provide clear indicators of customer discussion points, allowing businesses to swiftly comprehend common issues and emerging trends.

2.3. Key NLP Features Relevant

Modern NLP offers a powerful toolkit for interpreting customer feedback. The procedure begins with fundamental text cleaning - eliminating punctuation, special characters, and other noise that could distort analysis. Tokenisation then breaks the text into meaningful units, while normalisation ensures that words are handled consistently.
The real analytical power comes from techniques like n-gram analysis, which identifies frequently occurring phrases that reveal common customer concerns. Word frequency analysis and visualisation like word clouds provide immediate visual summaries of predominant themes. For this project, Latent Dirichlet Allocation (LDA) is the centrepiece analytical method, identifying hidden topics that structure customer conversations without any pre-defined categories. While sentiment analysis plays a secondary role here, it provides valuable context about the emotional tone surrounding different topics.

3. PROPOSED NLP AI SOLUTION ARCHITECTURE

3.1. Python Tools and Libraries

For text analytics, the proposed NLP solution leverages Python's ecosystem of data science libraries. Utilising "pandas" for structured data manipulation and "numpy" for numerical computations. The "re" module offers powerful capabilities to clean and normalise unstructured text data for text-specific preprocessing
The Natural Language Toolkit (nltk) provides essential text processing components for fundamental NLP functions, including stopword removal, advanced tokenisation, n-gram generation, and lemmatisation. These tools ensure the text data is properly standardised before analysis.
While visualisation with "matplotlib" and "seaborn" generates statistical plots to reveal trends, "wordcloud" provides intuitive representations of term frequencies. On the machine learning side, "scikit-learn" powers the analytical backbone, with "CountVectorizer" converting text into a numerical format and "LatentDirichletAllocation" facilitating unsupervised topic modeling. Custom implementations further enhance temporal analysis, allowing the tracking of topic evolution over time

3.2. Overview of the NLP Pipeline Architecture

3.2.1. Data Collection and Loading

Using Python's pandas library, we load a CSV file named 'reviews_data.csv' into a DataFrame structure. This initial dataset contains 850 rows across 6 different columns, capturing diverse aspects of customer feedback, including review text, ratings, timestamps, and other metadata

	name	location	Date	Rating	Review	Image_Links
0	Helen	Wichita Falls, TX	Reviewed Sept. 13, 2023	5.0	Amber and LaDonna at the Starbucks on Southwes...	['No Images']
1	Courtney	Apopka, FL	Reviewed July 16, 2023	5.0	** at the Starbucks by the fire station on 436...	['No Images']
2	Daynelle	Cranberry Twp, PA	Reviewed July 5, 2023	5.0	I just wanted to go out of my way to recognize...	['https://media.consumeraffairs.com/files/cach...
3	Taylor	Seattle, WA	Reviewed May 26, 2023	5.0	Me and my friend were at Starbucks and my card...	['No Images']
4	Tenessa	Gresham, OR	Reviewed Jan. 22, 2023	5.0	I'm on this kick of drinking 5 cups of warm wa...	['https://media.consumeraffairs.com/files/cach...
...	...	...	...	...	...	...
845	Becky	Agoura Hills, CA	Reviewed July 13, 2006	NaN	I ordered two venti frappacino's without whipp...	['No Images']
846	Bob	Goodrich, MI	Reviewed Jan. 3, 2005	NaN	No Review Text	['No Images']
847	Erik	Valley Village, CA	Reviewed Nov. 5, 2004	NaN	DEMANDED TIPS FROM ME, THEN MADE ME WAIT UNTIL...	['No Images']
848	Andrew	Fallbrook, CA	Reviewed Oct. 20, 2004	NaN	No Review Text	['No Images']
849	Christian	Ramsey, NJ	Reviewed July 19, 2000	NaN	No Review Text	['No Images']

850 rows × 6 columns

Figure 1. Raw dataset after importing

3.2.2. Data Exploration

The exploration reveals important findings about the data structure and quality. Examining the dataset shape (using df.shape() function) confirms we're working with 850 individual reviews with 6 different attributes. Additionally, checking data types (df.types() function) shows most columns are stored as text (object type), with only the rating column formatted as numerical data (float). We also identify that the Date column, currently stored as text, will need conversion for temporal analysis.

The data quality assessment uncovers 145 missing values in the rating column by applying isnull().sum() function, which will need to be addressed in subsequent data preparation steps. Basic statistical analysis provides insights into the distribution of scores, including measures of central tendency and dispersion (by using the describe() function) of the dataset.

```
Shape of the dataset: (850, 6)

Data types of each column:
name            object
location        object
Date            object
Rating         float64
Review          object
Image_Links     object
dtype: object

Missing values in each column:
name             0
location         0
Date             0
Rating         145
Review           0
Image_Links      0
dtype: int64

Descriptive statistics:
          name      location                  Date      Rating
count      850           850                   850  705.000000
unique     604           633                   741         NaN
top      Linda  New York, NY  Reviewed Sept. 14, 2017         NaN
freq        13            14                     4         NaN
mean       NaN           NaN                   NaN    1.870922
std        NaN           NaN                   NaN    1.397672
min        NaN           NaN                   NaN    1.000000
25%        NaN           NaN                   NaN    1.000000
```

Figure 2. Dataset Exploration output

3.2.3. Data Preparation

The data preparation phase transforms the raw dataset into a clean format through several systematic steps. Using pandas' dropna() function, we remove all rows with missing ratings that were identified during exploration, ensuring our analysis works only with complete records. This operation reduces our dataset from 850 to 705 entries while maintaining data integrity.

Furthermore, the Date column undergoes conversion to datetime format using pd.to_datetime(), allowing for temporal analysis and time-series visualisation. We then streamline our dataset by selecting only the most relevant columns for our NLP analysis - "Review", "Rating", and "Date" –

creating a new DataFrame (df1) that focuses on these key features while discarding less relevant metadata like names, locations, and image links.
To enhance the analytical capabilities, a new "Sentiment" feature is added to categorise reviews based on numerical ratings. Applying a simple yet effective categorisation scheme, we label reviews with ratings ≥4 as "positive", those with exactly 3 stars as "neutral", and ratings ≤ 2 as "negative". The final prepared dataset contains 705 complete records across 4 columns (Review, Rating, Date, and Sentiment), forming a solid foundation for subsequent text analysis and modeling.

	Review	Rating	Date	Sentiment
0	Amber and LaDonna at the Starbucks on Southwes...	5	2023-09-13	positive
1	** at the Starbucks by the fire station on 436...	5	2023-07-16	positive
2	I just wanted to go out of my way to recognize...	5	2023-07-05	positive
3	Me and my friend were at Starbucks and my card...	5	2023-05-26	positive
4	I'm on this kick of drinking 5 cups of warm wa...	5	2023-01-22	positive
...	...	...	...	...
700	I ordered Via Starbucks coffee online. I recei...	1	2011-10-02	negative
701	My name is Ric **, I am journalist by professi...	3	2011-08-31	neutral
702	The bagel was ice cold, not cut and not toasted.	1	2011-08-24	negative
703	In the morning of Monday, August 15, 2011, at ...	1	2011-08-15	negative
749	I found the coffee at Starbucks overrated and ...	5	2010-02-06	positive

705 rows × 4 columns

Figure 3. Prepared dataset

3.2.4. Exploratory Data Analysis (EDA)

The Exploratory Data Analysis phase provides crucial insights into the fundamental characteristics of our customer review dataset before proceeding with advanced modeling (Harinakshi et al., 2022). By examining the distribution of star ratings through a visually intuitive bar chart that employs a colour gradient from red (1-star) to green (5-star), figure 4 reveals a significant imbalance in the data, with 1-star ratings dominate with over 450 reviews and the aggregated negative reviews (1-2 stars) surpass positive ones (4-5 stars) by over fourfold, indicating a concerning pattern that warrants Starbucks' immediate attention to customer satisfaction issues.

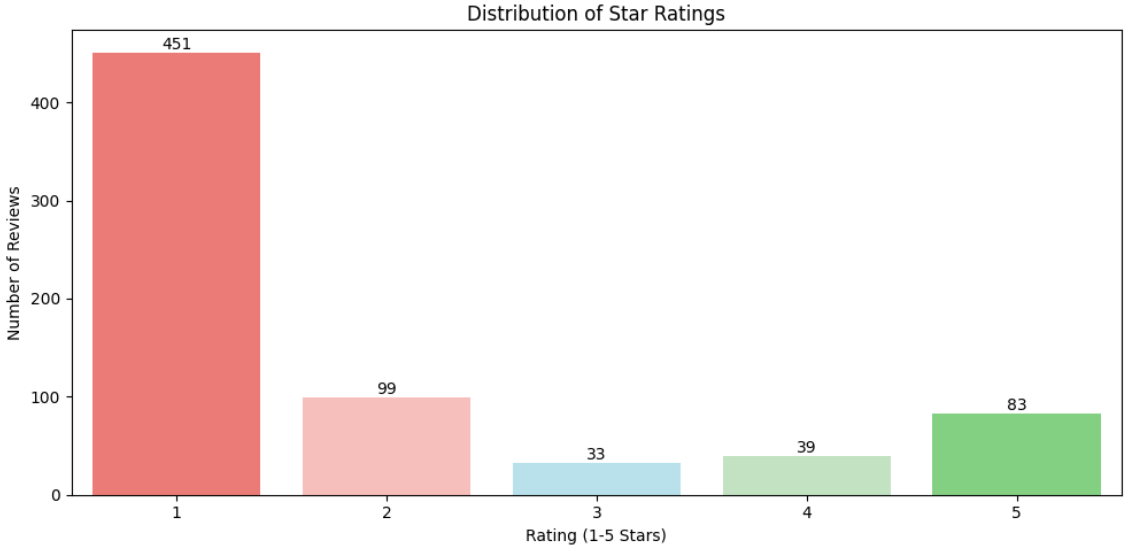

Figure 4. Distribution of Ratings

Sentiment distribution (Figure 5) shows negative sentiment accounts for 78% of all reviews, establishing it as the predominant tone across the dataset. Temporal analysis (Figure 6) demonstrates how this negativity has persisted over time, peaking in 2015 with 82 negative reviews due to controversial marketing campaigns and operational challenges (Mainwaring, 2015).

However, a well-received "A Good Year" campaign in 2017 improved sentiment (Cohen, 2017), suggesting potential strategic communication's impact on customer perception.

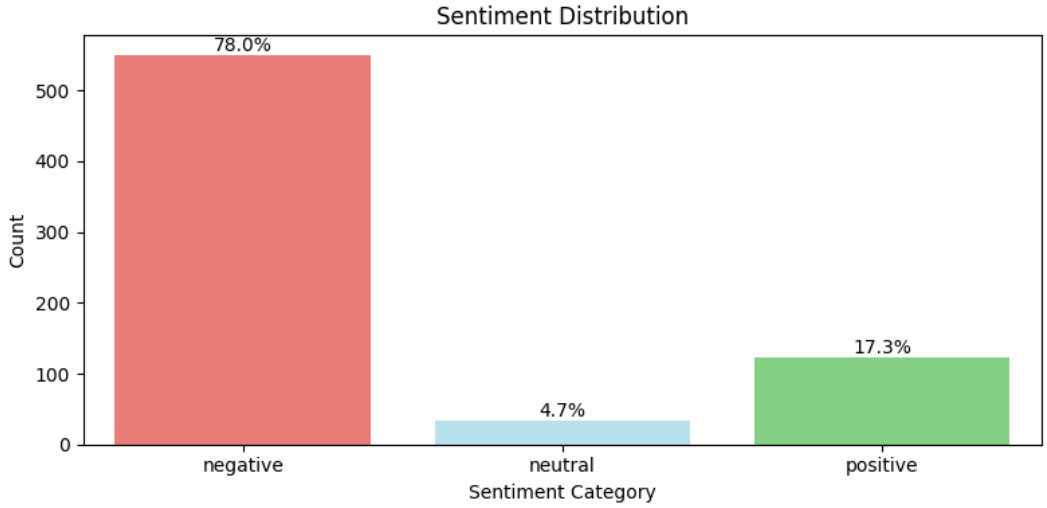

Figure 5. Distribution of Sentiment

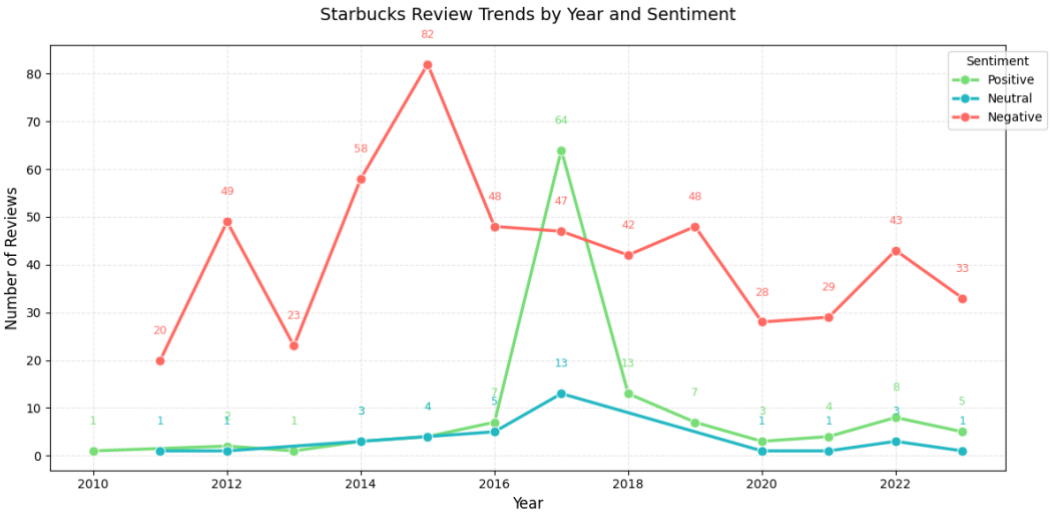

Figure 6. Line graph of Review Trends by Year and Senitment

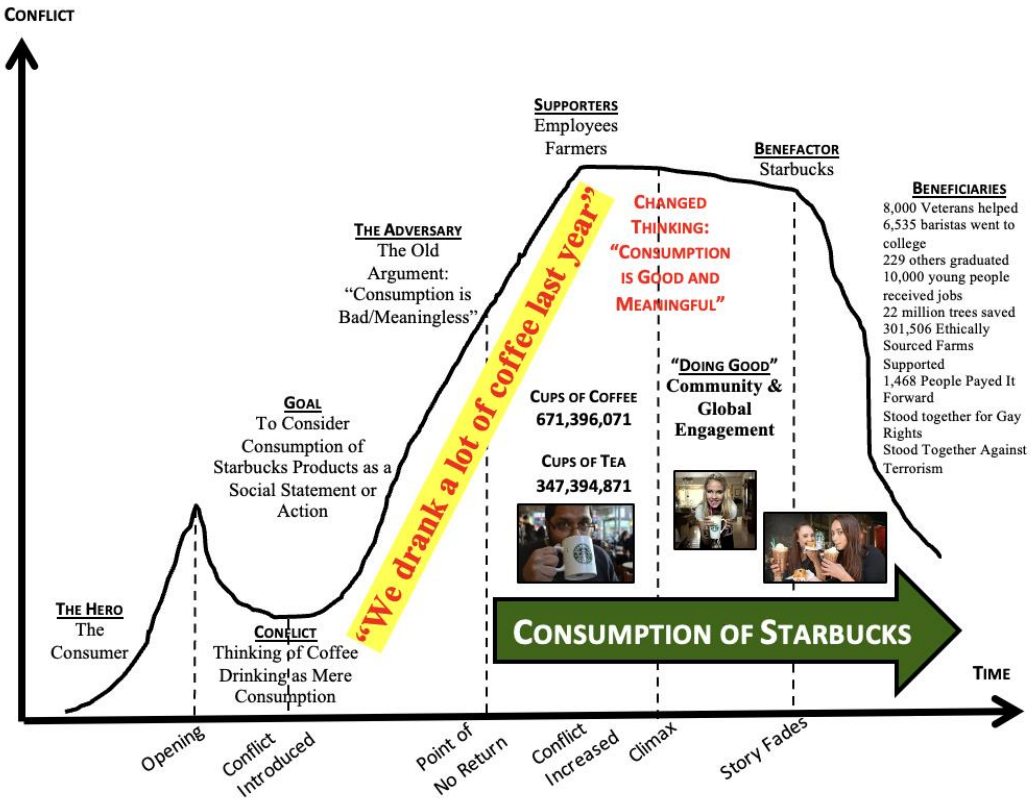

Figure 7. Starbucks' "A Good Year" Story Arc (Cohen, 2017)

A particularly revealing finding emerges from our analysis of review length patterns. By calculating character and word counts for each review and comparing them across sentiment categories, we discover that negative reviews tend to be significantly longer than positive ones. Summary statistics and boxplot visualisations (Figure 8) clearly demonstrate this relationship, supporting the psychological premise that dissatisfied customers often provide more detailed feedback as a form of catharsis or complaint justification. This length-sentiment correlation has crucial implications for our subsequent text analysis, as the more verbose negative reviews may contain richer thematic content for our topic modeling.

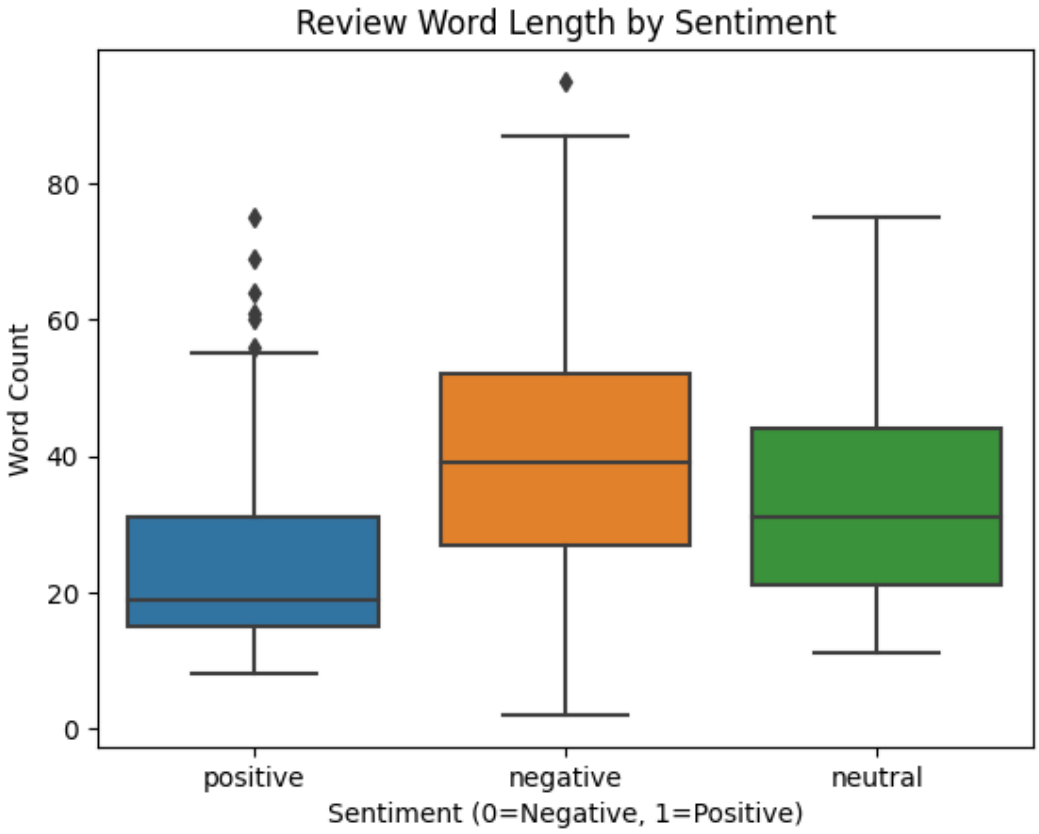

Figure 8. Boxplot of Word Length by Sentiment

3.2.5. Data Preprocessing (Text-Specific Cleaning)

The text preprocessing stage transforms raw customer reviews into a clean, analysis-ready format through a series of systematic NLP techniques. Begin by standardising text case, converting all characters to lowercase to ensure consistent treatment of words regardless of their original capitalisation (e.g., "Coffee" becomes "coffee"). This prevents the same word from being counted as distinct tokens based solely on case differences.

Additionally, comprehensive text cleaning uses regular expressions to remove various forms of noise. This includes eliminating punctuation marks, special characters, URLs, social media mentions, numerical values, and excess whitespace due to the reason that they typically don't contribute meaningful semantic content for topic modelling and could potentially distort analysis. The cleaned text is stored in a new column named "Cleaned_Review" while preserving the original review text for reference.

A critical step involves stopword removal, where we filter out common function words ("a", "an", "the") that appear frequently but carry little thematic meaning. Recognising the Starbucks-specific context of our data, we extend this to include brand-related terms like "Starbucks" and "coffee" that appear ubiquitously across reviews but don't help distinguish between different topics. This customisation prevents these high-frequency brand terms from dominating our topic models while maintaining truly meaningful content words.

Applying lemmatisation to reduce words to their base or dictionary forms (lemmas) and ensure accurate transformations. This process converts different inflected forms of a word (e.g., "drink", "drinks") to their common base ("drink"), significantly improving the model's ability to recognise semantic relationships while reducing vocabulary size. The final preprocessing step involves tokenisation, where we split each review into individual words (tokens), creating the fundamental units that our LDA model will analyse. This transformation converts each review from a continuous string into a sequence of meaningful tokens, preparing the data for vectorisation and subsequent topic modeling.

	Review	Rating	Date	Sentiment	Year	Cleaned_Reviews	Tokens	Char_Length	Word_Length
0	Amber and LaDonna at the Starbucks on Southwes...	5	2023-09-13	positive	2023	amber ladonna southwest parkway warm welcoming...	[amber, ladonna, southwest, parkway, warm, wel...	331	19
1	** at the Starbucks by the fire station on 436...	5	2023-07-16	positive	2023	fire station altamonte spring finally helped f...	[fire, station, altamonte, spring, finally, he...	555	41
2	I just wanted to go out of my way to recognize...	5	2023-07-05	positive	2023	wanted recognize employee billy franklin park ...	[wanted, recognize, employee, billy, franklin,...	351	30
3	Me and my friend were at Starbucks and my card...	5	2023-05-26	positive	2023	friend card work thankful worker paid drink ni...	[friend, card, work, thankful, worker, paid, d...	441	28
4	I'm on this kick of drinking 5 cups of warm wa...	5	2023-01-22	positive	2023	kick drinking cup warm water work instacart ri...	[kick, drinking, cup, warm, water, work, insta...	405	29
...	...	...	...	...	...	...	...	...	...
700	I ordered Via Starbucks coffee online. I recei...	1	2011-10-02	negative	2011	ordered via online received email stock wareho...	[ordered, via, online, received, email, stock,...	199	15
701	My name is Ric **, I am journalist by professi...	3	2011-08-31	neutral	2011	name ric journalist profession sending letter ...	[name, ric, journalist, profession, sending, l...	310	26
702	The bagel was ice cold, not cut and not toasted.	1	2011-08-24	negative	2011	bagel ice cold cut toasted	[bagel, ice, cold, cut, toasted]	48	5
703	In the morning of Monday, August 15, 2011, at ...	1	2011-08-15	negative	2011	morning monday august coworker stopped buy usu...	[morning, monday, august, coworker, stopped, b...	297	23
749	I found the coffee at Starbucks overrated and ...	5	2010-02-06	positive	2010	found overrated taste survey bitter comment sp...	[found, overrated, taste, survey, bitter, comm...	330	26

Figure 9. Analysis-ready dataset

3.2.6. Data Visualisation

The visualisation phase converts processed text data into intuitive graphical representations that reveal key patterns in customer feedback (Bashri & Kusumaningrum, 2017). Word frequency distributions across different sentiment categories (positive, neutral, negative) are analysed using a custom function (get_word_freq()). It processes tokenised reviews for each sentiment group and returns frequency counts through Python's Counter object, enabling us to determine which words are most commonly used in each type of review.

```
# Calculate Word Frequency by Sentiment
def get_word_freq(tokens_list):
    all_words = [word for tokens in tokens_list for word in tokens]
    return Counter(all_words)

positive_freq = get_word_freq(df1[df1['Sentiment'] == 'positive']['Tokens'])
neutral_freq = get_word_freq(df1[df1['Sentiment'] == 'neutral']['Tokens'])
negative_freq = get_word_freq(df1[df1['Sentiment'] == 'negative']['Tokens'])
```

Figure 10. The code working for Calculate Word Frequency by Sentiment

3.2.6.1. Word Cloud

Word clouds serve as our primary visualisation tool for displaying overall term frequencies in an understandable manner. Using the WordCloud library, we generate sentiment-specific word clouds with colour-coding: green for positive, blue for neutral, and red for negative reviews. The size of each word in these clouds corresponds to its frequency in the corresponding sentiment category, allowing for immediate visual comparison of dominant terms across different types of feedback. While effective for showing relative term prominence, we recognise word clouds' limitations in conveying exact frequency counts or sentiment polarity.

3.2.6.2. Frequency Plot

For more precise quantitative analysis, frequency plots are created that display word counts in structured bar charts. The top 20 most frequent words across all reviews provide managers with a snapshot of dominant themes. More insightful is the clustered bar chart comparing word frequencies between positive and negative reviews, which highlights terms that disproportionately appear in either category. For instance, words like "money" appear more frequently in negative reviews, potentially indicating pricing concerns

Temporal analysis through yearly word frequency tracking reveals how customer concerns' evolution. By grouping reviews by year and plotting the top terms for each period, we can identify emerging issues or fading concerns.

3.2.6.3. N-Grams

We expand our analysis beyond single words to analyse significant word combinations through n-grams. Bigrams and trigrams capture more nuanced expressions that single words might miss, such as "long wait time" or "great customer service". These multi-word patterns provide richer context for understanding customer sentiment and identifying specific areas requiring attention. The n-gram frequency distributions reveal how words combine to form common expressions of praise or complaint in Starbucks reviews, enhancing other visualisation

3.2.7. Topic Modeling with LDA

3.2.7.1. LDA in general

Latent Dirichlet Allocation (LDA) is a key analytical technique for uncovering hidden themes within customer reviews. It operates through two key processes: a generative process that creates documents based on known topic distributions, and an inference process that works backward to discover these latent patterns in existing documents.

In our Starbucks analysis, LDA is employed in its inference capacity, using scikit-learn's implementation to reverse-engineer the most probable topics. The mathematical foundation calculates two key probability distributions: φ_k representing the probability of words within topics and θ_d representing the proportion of topics within documents, allowing for identifying coherent themes related to service quality, product preferences, or loyalty program experiences (Bashri & Kusumaningrum, 2017).

$$\varphi_k = p(w = t|z = k) = \frac{n_{t,k} + \beta_t}{\sum_{t=1}^{V} n_{t,k} + \beta_t}$$

$$\theta_d = p(z = k|d) = \frac{n_{d,k} + \alpha_k}{\sum_{k=1}^{K} n_{d,k} + \alpha_k}$$

3.2.7.2. Output Interpretation for Managers

The practical output of LDA provides managers with actionable insights through clearly interpretable topics. Each discovered theme manifests as a cluster of semantically related words - for instance, a topic containing terms like "reward," "free," and "program" would clearly indicate customer concerns about Loyalty Program. These topic signatures enable data-driven decision making by highlighting both strengths to leverage and pain points requiring intervention

3.2.7.3. View the Topic changing over times

Beyond static topic identification, we extend our analysis to track how these themes evolve temporally. By applying LDA separately to reviews from different periods, we can observe shifting customer priorities. This temporal dimension transforms our topic modeling from a diagnostic tool into a strategic early warning system, helping organisations anticipate rather than just react to customer experience trends.

3.3. Topic Modeling with LDA (detailed implementation)

Strategic random sampling of the analysed-ready dataset is the first step in our LDA implementation, which strikes a balance between processing efficiency and analytical rigour. By setting a fixed random_state parameter, we ensure reproducible results while working with a manageable portion of data. The sample size is carefully chosen to maintain representative coverage of customer feedback patterns while enabling faster model iteration.

The text vectorization process employs CountVectorizer to transform preprocessed reviews into a document-term matrix (DTM), which aims to convert textual data into a numerical format, while incorporating several optimisations. We configure the vectorizer to filter out overly common and rare terms through min_df and max_df parameters, apply our custom stopword list containing Starbucks-

specific terms, and limit the vocabulary size to focus on the most meaningful features. The fit_transform() method efficiently processes all reviews to create our input matrix

The LDA model uses scikit-learn's LatentDirichletAllocation with carefully tuned parameters, configured with 10 topics. The "Online" learning method is selected for its computational efficiency with larger datasets. The model is trained on the DTM using the fit() method, iteratively discovers the latent topic structure in customer feedback.

```python
# Random sampling (adjust sample_size as needed)
sample_size = 700  # Typical for mid-sized datasets
sampled_df = df1.sample(n=sample_size, random_state=42)  # Reproducible sampling

vectorizer = CountVectorizer(
    max_df=0.95,      # Ignore overly common words
    min_df=5,         # Require words to appear in ≥5 reviews
    stop_words='english',
    max_features=8000  # Limit vocabulary size
)

# Add custom stopwords
vectorizer.stop_words_ = vectorizer.get_stop_words().union(set(my_stopwords))

# Create document-term matrix
dtm = vectorizer.fit_transform(sampled_df['Cleaned_Reviews'])
print(f"DTM shape: {dtm.shape}")  # Should be (sample_size, vocab_size)
lda_model = LatentDirichletAllocation(
    n_components=10,           # Start with 10 topics
    random_state=42,
    learning_method='online',  # Better for larger datasets
    max_iter=15,               # More iterations for stability
    batch_size=256             # Faster training
)

lda_model.fit(dtm)
# Print topics with top 10 words
feature_names = vectorizer.get_feature_names_out()

for topic_idx, topic in enumerate(lda_model.components_):
    top_features_ind = topic.argsort()[:-10 - 1:-1]  # Top 10 word indices
    top_features = [feature_names[i] for i in top_features_ind]
    print(f"Topic {topic_idx}: {', '.join(top_features)}")
```

Figure 11. LDA Model Code working

```
DTM shape: (700, 886)
Topic 0: card, customer, account, service, called, gift, email, money, store, order
Topic 1: star, reward, purchase, time, use, free, program, mask, bonus, offer
Topic 2: customer, service, time, drink, store, food, ive, like, great, cup
Topic 3: milk, order, time, ordered, customer, cup, tea, latte, grande, make
Topic 4: order, caramel, time, location, long, wait, drive, mocha, drink, ordered
Topic 5: customer, store, like, time, employee, manager, location, year, service, people
Topic 6: sandwich, cheese, food, man, delicious, enter, employee, old, usually, serving
Topic 7: ice, extra, gave, cup, tip, manager, cold, water, ordered, hot
Topic 8: store, cup, order, minute, hot, ordered, waited, bag, blend, service
Topic 9: line, people, order, store, morning, waiting, open, person, counter, music
```

Figure 12. LDA Model Output

Following model training, we extract and examine the top words for each topic to facilitate interpretation. These word distributions form the foundation for our topic labeling process, where we assign meaningful, business-relevant names to each topic based on its characteristic terms. For instance, a topic showing high probabilities for words like "card", "account", and "gift" would be labeled as "Gift Card & Payment Issues". This translation from statistical output to business insight is critical for managerial usability.

The trained model is then applied to our full dataset, with each review assigned to its dominant topic through probability analysis. We store these assignments in a new column, enabling topic-based filtering and aggregation. Most importantly, temporal topic analysis tracks prevalence across different years, identifying growing or diminishing customer concerns

4. DATA PRESENTATION

4.1. Unigrams – Identifying Core Themes

Our unigram visualisations provide two complementary perspectives on customer feedback. The word cloud (Figure 13) offers an immediate, intuitive understanding of dominant themes through its size-based representation, where terms like "customer," "service," "store," and "time" appear most prominently. This visualisation excels at conveying overall patterns immediately, making it particularly valuable for initial exploratory analysis. The accompanying bar chart (Figure 14) delivers precise quantitative data, ranking the top 20 words by frequency to enable detailed comparison and measurement.

Figure 13. Overal Word frequency in Word Cloud

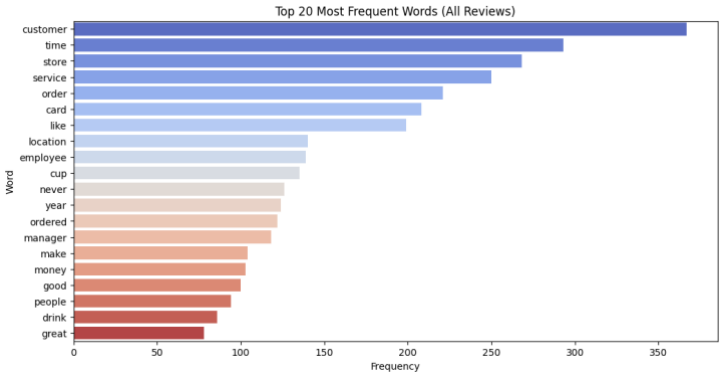

Figure 14. Overal Word frequency in Horizonal Bar Chart

When segmented by sentiment, these visualizations reveal important nuances. The positive sentiment word cloud (Figure 15) highlights terms like "service", suggesting aspects that satisfy customers, while the negative cloud (Figure 17) emphasizes words like "time," "order," and "never" that signal frustration points. The comparative bar chart further demonstrates that while most words appear in both positive and negative reviews, their frequencies differ significantly, supporting existing research that negative reviews tend to be more impactful and frequent (Kim, 2023). This asymmetry suggests customers are more motivated to share negative experiences, making these insights particularly valuable for service improvement.

Figure 15. Positive Wordcloud

Figure 16. Neutral Wordcloud

Figure 17. Negative Wordcloud

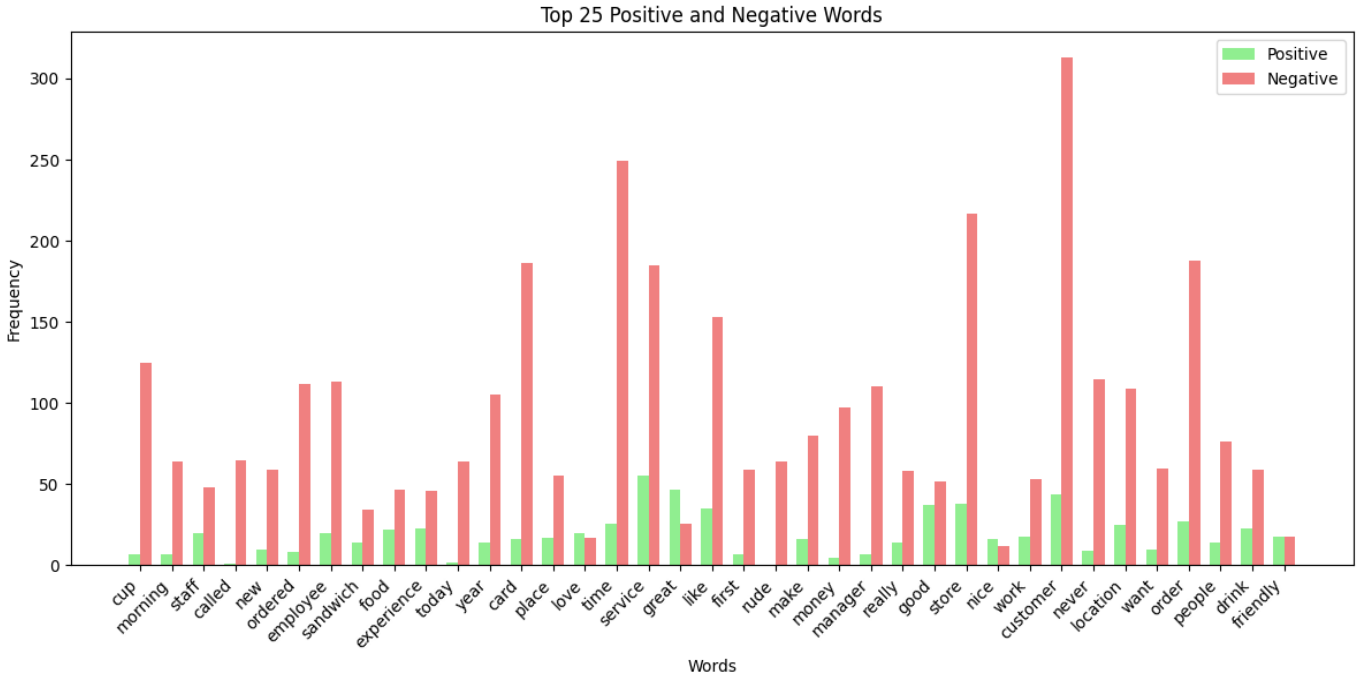

Figure 18. Clustered Bar chart: Word Distribution by Sentiment

4.2. N-Grams – Understanding Contextual Phrases

Moving beyond individual words, our bigram and trigram analysis uncovers the contextual phrases that reveal more specific customer experiences. The bigram visualisation (Figure 19) identifies key combinations like "customer service" and "caramel macchiato", highlighting both operational aspects and product-specific feedback. The trigram analysis proves even more revealing, with phrases like "worst customer service" appearing with notable frequency (approximately 12 occurrences). These multi-word patterns provide actionable specificity that single-word analysis cannot match - where a unigram might show "service" as common, the trigram pinpoints exactly how that service is being criticized.

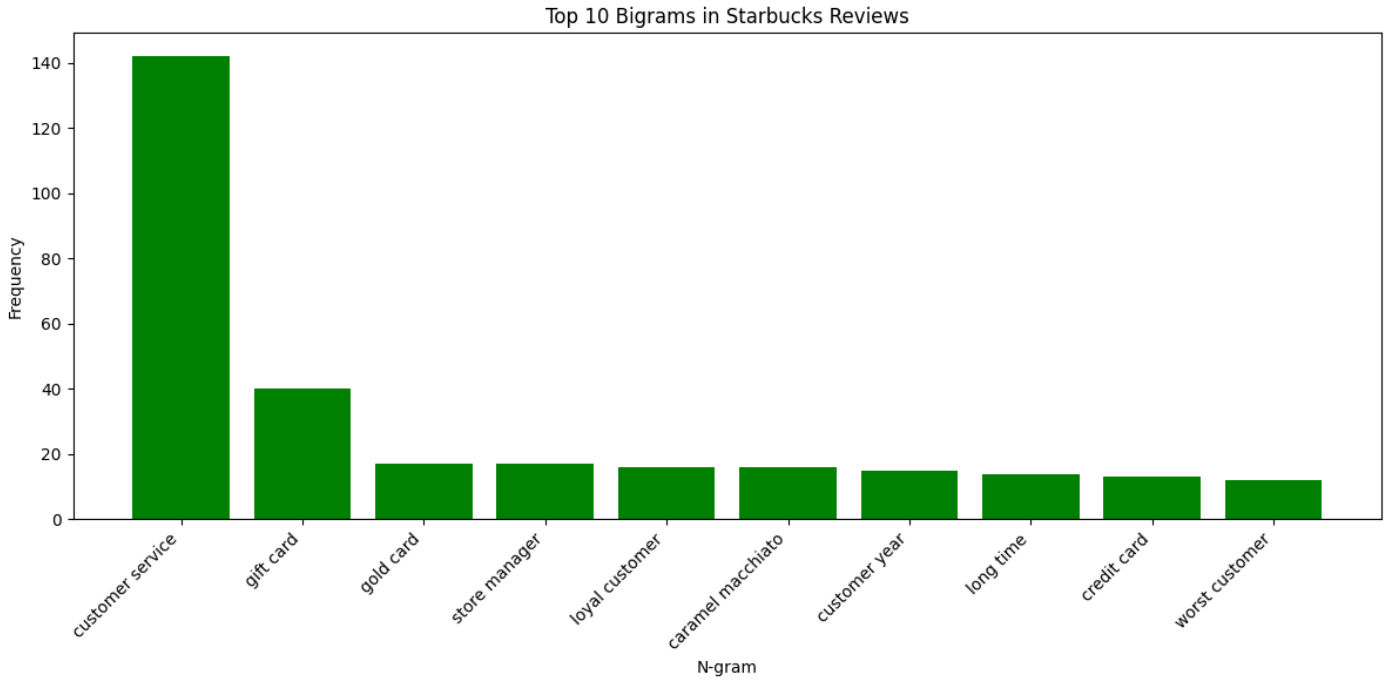

Figure 19. Bigrams distribution by Bar chart

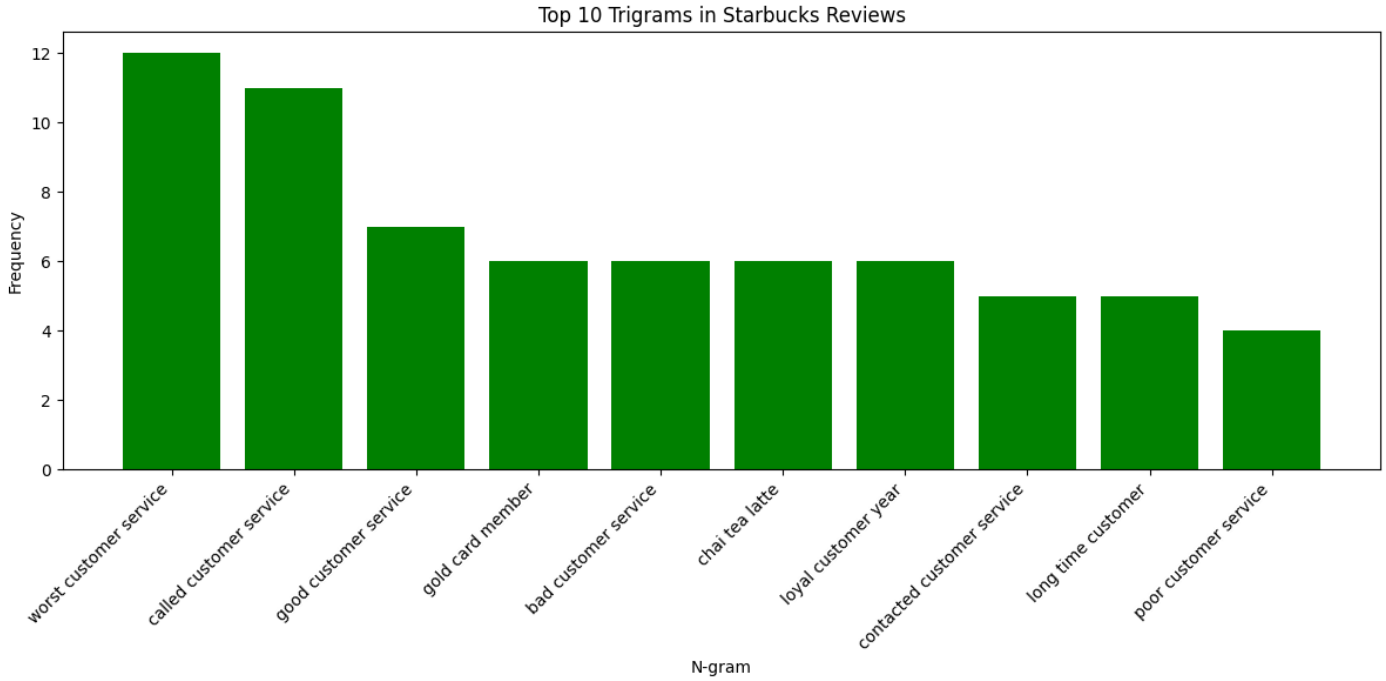

Figure 20. Trigrams distribution by Bar chart

4.3. LDA Topic Trends – Temporal Analysis of Key Issues

Our temporal topic analysis tracks how customer concerns evolve over the years 2010-2023. The line graph visualisation (Figure 21) reveals several important patterns: Customer Experience topics show a dramatic decline after 2010, suggesting operational improvements addressed early pain points. However, Store Staff & Management topics remain consistently prominent, indicating ongoing challenges in human resource management

These temporal patterns enable managers to validate past interventions, identify persistent issues, and spot emerging trends for strategic planning.

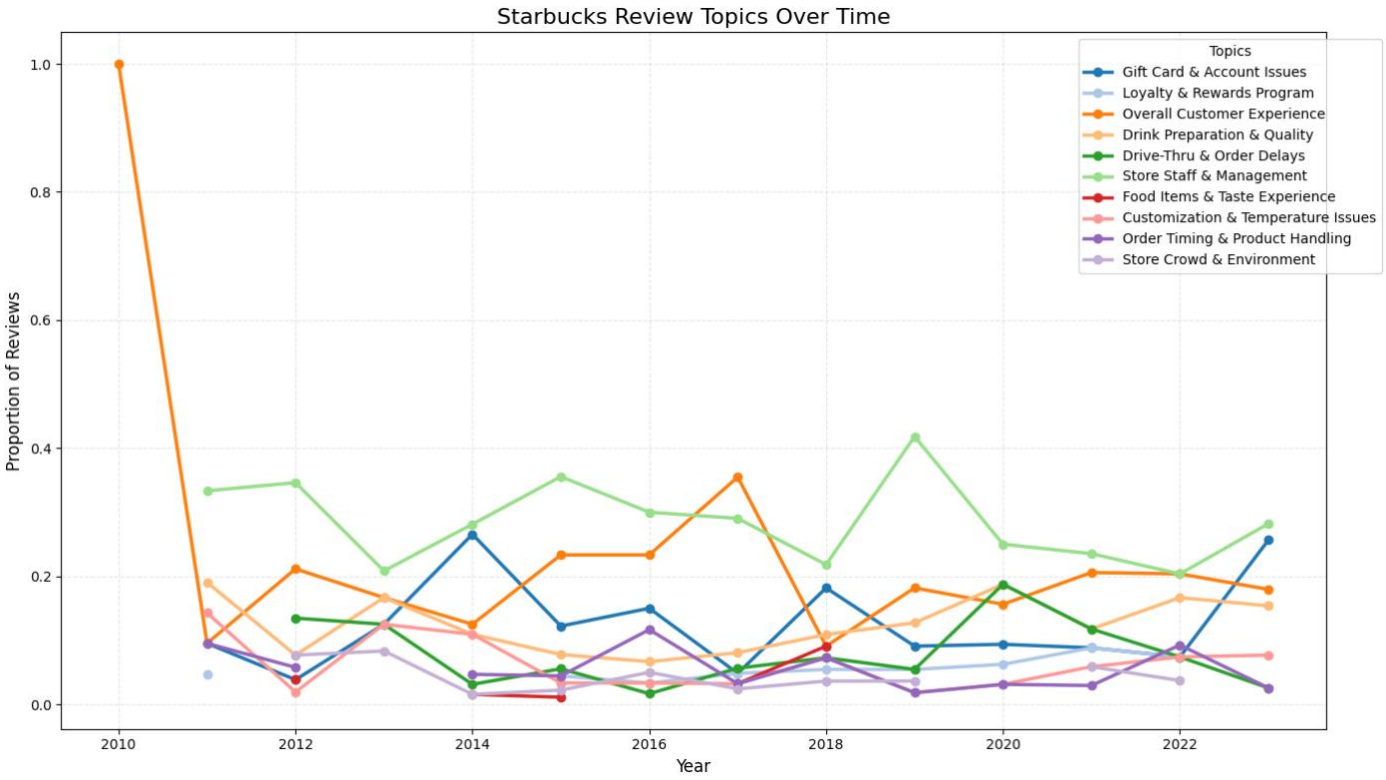

Figure 21. Top 10 Topic Modeling after LDA training

5. DISCUSSION AND CONCLUSION

5.1. Main points

The analysis addresses a critical challenge facing modern businesses - the overwhelming volume of unstructured customer feedback that traditional analytical methods cannot efficiently handle. Our NLP-powered solution, utilising LDA algorithm, effectively identifies hidden themes and patterns at scale. The visualisation reveals that customer concerns primarily cluster around customer experience, service quality, and wait times, with particular emphasis on staff behaviour and store management topics. These findings align with pain points in the retail food service industry, providing managers with a powerful tool for prioritising interventions and concentrating resources on potential areas that impact customer satisfaction.

5.2. Ethical consideration

While our analytical approach delivers significant business value, it also raises important ethical considerations that require careful attention. The use of customer feedback data, even when publicly available, necessitates thoughtful handling of privacy concerns. We recommend implementing strict data anonymisation protocols to remove all personally identifiable information, ensuring compliance with data protection regulations.

The potential for algorithmic bias presents another critical consideration. Our analysis may unintentionally increase existing biases in customer feedback, particularly in judgements of staff performance across different locations. To mitigate this risk, we suggest implementing bias detection system and combining automated analysis with human oversight when making sensitive personnel decisions.

Transparency in model outputs represents the third key ethical concern. The statistical nature of LDA means topics may occasionally require expert interpretation to avoid misapplication. Clear documentation of methodology and constraints, along with appropriate interpretation training, is recommended to ensure ethical deployment.

REFERENCE

Bashri, M. F. A., & Kusumaningrum, R. (2017). Sentiment analysis using Latent Dirichlet Allocation and topic polarity wordcloud visualization. 2017 5th International Conference on Information and Communication Technology (ICoIC7), 1–5. https://doi.org/10.1109/ICoICT.2017.8074651
Cohen, J. Consumption of Starbucks. 10.13140/RG.2.2.28834.48324

Flayeh, A. K., Hamodi, Y. I., & Zaki, N. D. (2022). Text Analysis Based on Natural Language Processing (NLP). 2022 2nd International Conference on Advances in Engineering Science and Technology (AEST), 774–778. https://doi.org/10.1109/AEST55805.2022.10413039

Harinakshi, Lydia, A. A., Poongundran, M., Masarath, S., Karthick, P. V., & Zayats, I. M. (2022). EDA and its Impact in Dataset Discover Patterns in the Service Sector. 2022 4th International Conference on Inventive Research in Computing Applications (ICIRCA), 940–945. https://doi.org/10.1109/ICIRCA54612.2022.9985599

Ittoo, A., Nguyen, L. M., & van den Bosch, A. (2016). Text analytics in industry: Challenges, desiderata and trends. Computers in Industry, 78, 96–107. https://doi.org/10.1016/j.compind.2015.12.001

Jelodar, H., Wang, Y., Yuan, C., Feng, X., Jiang, X., Li, Y., & Zhao, L. (2019). Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey. Multimedia Tools and Applications, 78(11), 15169–15211. https://doi.org/10.1007/s11042-018-6894-4

Kim, R. Y. (2023). Text Mining Online Reviews: What Makes a Helpful Online Review? IEEE Engineering Management Review, 51(4), 1–18. https://doi.org/10.1109/EMR.2023.3286349

Liu, C., Liu, K., & Ma, J. (2020). Research on Innovation of Enterprise Business Model Based on Big Data Analysis. 2020 International Conference on Big Data and Social Sciences (ICBDSS), 178–181. https://doi.org/10.1109/ICBDSS51270.2020.00047

Mainwaring, S. (2015). *Starbucks Finds Itself In Hot Water For Talking About Race*. [online] Forbes. Available at: https://www.forbes.com/sites/simonmainwaring/2015/03/23/starbucks-finds-itself-in-hot-water-for-talking-about-race/.

Mishra, P., Ninawe, K., Dhawle, K., Prasad, S., Band, A., & Dubey, P. (2024). NLP in Business Analytics: Generating Insights from Textual Data Using NLP Models. Proceedings (International Conference on Computational Intelligence and Communication Networks), 1472–1478. https://doi.org/10.1109/CICN63059.2024.10847530

Sintoris, K., & Vergidis, K. (2017). Extracting Business Process Models Using Natural Language Processing (NLP) Techniques. The Institute of Electrical and Electronics Engineers, Inc. (IEEE) Conference Proceedings, 1, 135-. https://ieeexplore.ieee.org/abstract/document/8010715