# Analysis of Customers' Reviews

Flavio Kaci Francesco Lollato Alex Meggiolaro

## 1. Introduction

Our project aims to conduct an analysis of reviews concerning McDonald's fast food restaurants, with the objective of identifying and understanding the main issues related to their service. To achieve this, we will utilize a combination of advanced machine learning techniques: Sentiment Analysis through Recurrent Neural Networks (RNN) and Topic Modeling via Latent Dirichlet Allocation (LDA). Sentiment Analysis will allow us to classify each review based on the expressed sentiment, categorizing them as positive, negative, or neutral. This classification provides a clear picture of the overall customer satisfaction. Meanwhile, Topic Modeling will help us uncover recurring themes and specific issues mentioned across the negative reviews, giving us insights into common areas of concern. By integrating these two methodologies, we can not only gauge the general sentiment of customer feedback but also pinpoint specific problems consistently reported by customers. This comprehensive analysis will enable McDonald's to better understand their customers' experiences and take targeted actions to improve their service quality.

## 2. Dataset

The dataset contains over 33,000 anonymized reviews of McDonald's stores in the United States, scraped from Google reviews. It provides valuable insights into customer experiences and opinions about various McDonald's locations across the country. The dataset includes information such as store names, categories, addresses, geographic coordinates, review ratings, review texts, and timestamps.

For our purposes, we have decided to focus solely on the content of the reviews and the ratings provided by customers. Therefore, we will eliminate all variables except **review** and **rating**. Other variables will be excluded from the analysis as they are not directly necessary for our analysis. This approach enables us to concentrate on the core feedback from customers and gain a clearer understanding of their experiences and opinions through the reviews and ratings.

## 3. General Data Pre-processing

After removing the not relevant variables, we proceeded with applying some basic pre-processing techniques to prepare our data to be modelled:

- **Normalization:** all reviews are converted to lowercase to ensure uniformity.
- **Tokenization:** each review is converted from a set of symbols to a sequence of word/tokens.
- **Lemmatization:** words are reduced to their base form to group different inflections of the same word.
- **Stopwords Removal:** common words and punctuation are removed, along with additional uninformative terms identified during review.

## 4. Sentiment Analysis

Sentiment analysis is a branch of natural language processing (NLP) that identifies and categorizes opinions in text as positive, negative, or neutral. It has become a key tool for understanding public opinion and customer feedback, particularly with the rise of user-generated content on platforms like social media and online reviews.

Using machine learning techniques, sentiment analysis classifies text based on its emotional tone. It has applications in various fields, such as marketing, finance, politics, and customer service.

In this project, we will apply sentiment analysis using two recurrent neural network (RNN) models. Our aim is to analyze customer reviews and feedback to predict the underlying emotional tone of the text and gain insights into criticisms and opinions related to McDonald's fast food restaurants. By leveraging these models, we hope to better understand customer sentiment, identifying areas of satisfaction and points of concern regarding McDonald's customer experience.

### 4.1. Pre-processing for Neural Network models

After converting the rating values to integers and then classifying sentiments, we applied several pre-processing techniques to prepare our data for modeling:

- **Conversion of Ratings to Numeric Values:** Ratings are converted from string format to integer values for easier numerical processing.
- **Sentiment Classification:** Ratings are mapped to sentiment categories such as 'positive', 'neutral', or 'negative' based on their numerical values.
- **Tokenization:** Reviews are transformed from raw text into sequences of integer tokens using a tokenizer that maps words to unique indices.
- **Sequence Padding:** Sequences of tokens are padded to ensure uniform input length across all reviews, making them suitable for input into an RNN model.
- **Conversion of Sentiment Labels to Numerical Format:** Sentiment categories are converted into numerical labels, which are then one-hot encoded for classification tasks.
- **Data Splitting:** The dataset is divided into training and test sets to evaluate model performance.

### 4.2. Simple Recurrent Neural Network (RNN)

Recurrent Neural Networks (RNNs) handle sequential data by maintaining a hidden state that captures information from previous time steps. They are effective for tasks like time series analysis and natural language processing but struggle with long-term dependencies due to the vanishing gradient problem.

Core components of RNNs include:

- **Hidden state**: A dynamic memory updated at each time step to retain past information.
- **Recurrent connections**: Connections that loop back, passing information across time steps.
- **Activation function**: Functions like tanh introduce non-linearity to learn complex patterns.

While RNNs capture short-term dependencies, they often struggle with long-term sequences, making advanced architectures like Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU) preferable for such tasks.

### 4.3. Long Short-Term Memory (LSTM)

Long Short-Term Memory (LSTM) networks are a type of Recurrent Neural Network (RNN) that are particularly well-suited for learning long-term dependencies in sequential data. Unlike traditional RNNs, LSTMs are explicitly designed to overcome the limitations of standard RNN architectures, particularly the vanishing gradient problem, which can occur when learning over long sequences of data.

LSTM networks are composed of several key components that work together to regulate and maintain the flow of information over time:

- **Cells**: These form the core of the LSTM architecture, responsible for maintaining the cell state, which acts as a memory across time steps.
- **Gates**: LSTMs use three gates to control the flow of information:
  - *Input gate*: Determines which values from the input should be updated in the cell state.
  - *Forget gate*: Decides which information should be discarded from the cell state.
  - *Output gate*: Controls which parts of the cell state are output at each time step.

| Layer (type) |
| --- |
| input_layer (InputLayer) |
| embedding (Embedding) |
| simple_rnn (SimpleRNN) |
| dropout (Dropout) |
| simple_rnn_1 (SimpleRNN) |
| dropout_1 (Dropout) |
| dense (Dense) |
| dense_1 (Dense) |

Figure 1. RNN: Model Architecture

| Layer (type) |
| --- |
| input_layer_1 (InputLayer) |
| embedding_1 (Embedding) |
| lstm (LSTM) |
| dropout_2 (Dropout) |
| lstm_1 (LSTM) |
| dropout_3 (Dropout) |
| dense_2 (Dense) |
| dense_3 (Dense) |

Figure 2. LSTM: Model Architecture

### 4.4. Results and comparison

For the sentiment analysis, we use RNN and LSTM models with the architectures described in Figures 1 and 2, respectively. These figures illustrate the specific configurations and components of the models used in this analysis. To compare the test set results of the two models, we analyze the confusion matrices, which provide a more detailed assessment of the performance of the RNN and LSTM models in sentiment classification.

In the case of the RNN, the model shows a good ability to correctly identify positive sentiments, with a high number of true positives. However, there is significant confusion regarding negative and neutral sentiments. In particular, the neutral class represents a notable challenge for the model, with many neutral examples incorrectly classified as positive or negative. This results in a high number of false negatives in the neutral class, while the positive class is widely overestimated. The confusion between classes is also evident from the model's tendency to classify too many negative examples as positive, creating a significant number of false negatives in the negative class. These observations are supported by the RNN's performance metrics: accuracy of 70.28%, precision of 0.74, recall of 0.70, and F1 score of 0.68.

Turning to the LSTM, overall performance improves compared to the RNN, especially in classifying negative

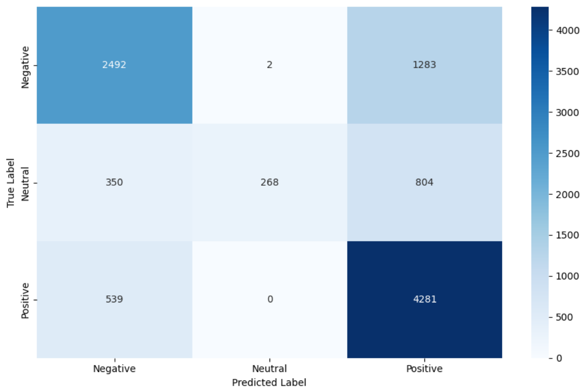


Figure 3. RNN: Confusion Matrix

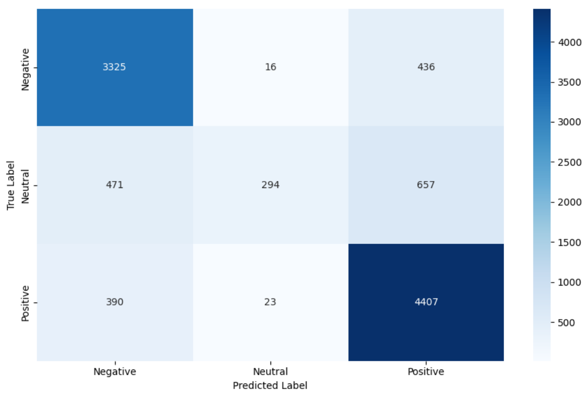


Figure 4. LSTM: Confusion Matrix

sentiments. The LSTM model reduces the number of false negatives in the negative class and slightly improves the recognition of neutral sentiments. However, as with the RNN, the neutral class continues to pose a difficulty, with a significant number of neutral examples being incorrectly classified as positive and negative. The LSTM also overestimates positive sentiments, but to a lesser extent compared to the RNN, showing a more balanced distribution between the classes. The performance metrics for the LSTM indicate an improvement over the RNN, obtaining an accuracy of $80.11\%$, precision of $0.81$, recall of $0.80$, and F1 score of $0.77$.

In summary, both models share a weakness in accurately recognizing neutral sentiments, with many examples being confused with extreme classes (negative or positive). This trend is more pronounced in the RNN compared to the LSTM, which, however, does not fully resolve the issue. The LSTM offers better performance in classifying negative sentiments, but like the RNN, it has a slight tendency to overestimate the positive class. Overall, the LSTM manages the classes better than the RNN, but both models could benefit from further improvements in recognizing the neutral class, which remains particularly problematic.

| Metric | RNN | LSTM |
|---|---|---|
| Accuracy (%) | 70.28 | 80.11 |
| Precision | 0.74 | 0.81 |
| Recall | 0.70 | 0.80 |
| F1 score | 0.68 | 0.77 |

Table 1. Comparison between RNN and LSTM evaluation metrics.

## 5. Topic Modelling

After using the LSTM network to perform sentiment analysis on customer reviews, we identified negative reviews as a valuable source of information for gaining deeper insights into specific areas of customer dissatisfaction. The predictions from the sentiment analysis allowed us to isolate these negative reviews, which we now analyze in more detail using topic modelling.

Topic Modeling is a powerful technique used to uncover hidden themes and patterns within a large collection of texts. By applying Topic Modeling to reviews of McDonald's fast food restaurants, we can identify the main topics that customers frequently discuss. One of the most effective methods for this task is Latent Dirichlet Allocation (LDA), which statistically analyzes the words in the reviews to find groups of words that tend to occur together. These groups, or topics, reveal underlying issues and common themes. This insight helps McDonald's understand the key areas of concern from their customers' perspectives, enabling them to address specific problems and improve their service.

### 5.1. LDA: Quick Overview

Latent Dirichlet allocation (LDA) is a generative probabilistic model for collections of discrete data such as text corpora. The basic idea behind this Bayesian model is that documents are represented as random mixtures over latent topics, where each topic is characterized by a distribution over words. Specifically, LDA assumes the following generative process for each document in the corpus:

- Choose $\theta \sim Dir(\alpha)$: the topic distribution for document d.
- For each of the N words $w_n$:
  - Choose a topic $z_n \sim Multinomial(\theta)$.
  - Choose a word $w_n$ from $p(w_n|z_n;\beta)$, a multinomial probability conditioned on the topic $z_n$.

$Alpha$ and $Beta$ are Dirichlet prior concentration parameters that represent respectively the document-topic density and the topic-word density.

### 5.2. Pre-Processing for LDA

The following steps outline the pre-processing pipeline applied to the negative reviews dataset, which was obtained from the sentiment analysis model:

- **POS Tagging:** Part-of-Speech (POS) tagging is applied to identify nouns and adjectives, which are crucial for topic modeling.
- **Noun and Adjective Filtering:** Only nouns and adjectives are retained, as they carry the most significant meaning for topic identification.
- **Frequent and Rare Terms Removal:** Words that occur too frequently (e.g., "order", "food") or too infrequently are discarded, as both types can introduce noise into the model.
- **Uninformative Terms Removal:** Other uninformative words (e.g., "mcdonald", "eat") are removed, since they do not contribute significantly to the semantic content of the text.

Now we are ready to apply LDA, but first we have to handle the choice of the number of topics, which is the main hyperparameter of LDA.

### 5.3. Ideal Number of Topics

In a Latent Dirichlet Allocation (LDA) model, the choice of the ideal number of topics is crucial to obtain meaningful and interpretable results. Too few topics can lead to too much overlap between topics, making it difficult to distinguish between different concepts and limiting the granularity of the analysis. Conversely, a too large number of topics can cause content fragmentation, with topics lacking coherence or clear meaning, making it difficult to interpret the results. In our project, in order to determine the most appropriate number of topics, we used the Coherence Score, a metric that measures how much the terms within each topic are semantically related, thus helping us to choose the number of topics that would ensure maximum consistency and interpretability of the results.

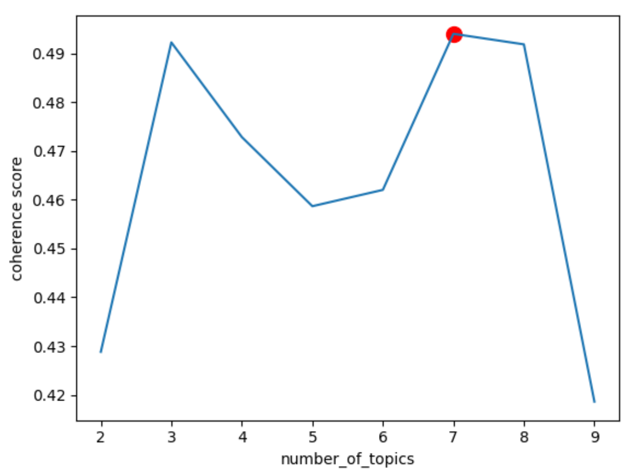


Figure 5. Ideal number of topics

It can be observed from Figure 5 that the ideal number of topics is 7 in terms of coherence score. However, after implementing the model, we observed that some topics were difficult to interpret, and there was significant content fragmentation between certain topics. As a result, we opted for a model with 3 topics. Although this led to a slight decrease in the coherence score, it resulted in a marked improvement in terms of interpretability.

### 5.4. LDA with 3 Topics: Results and Interpretation

In the following we present the table of the relevant words per topic:

| Waiting time | Staff | Location and Food quality |
|---|---|---|
| time | customer | fry |
| minute | employee | bad |
| drive | manager | place |
| wrong | rude | chicken |
| line | people | table |
| window | staff | soda |
| wait | lady | dirty |
| slow | worker | bathroom |
| car | chashier | cold |
| hour | attitude | sandwich |

Table 2. Relevant Words per Topic

The table 2 categorizes the words into the three distinct topics, which we have labeled as 'Waiting Time,' 'Staff,' and 'Location and Food Quality' based on the content and context of the associated words.

- **Waiting Time**: This topic encompasses terms related to the duration of service or waiting periods experienced by customers. In addition, words like 'drive,' 'line,' 'window,' and 'car,' highlight concerns related to the slow service experienced at the "Mc-Drive". These terms suggest that this topic primarily captures feedback related to the speed of service.
- **Staff**: The second topic focuses on various aspects of staff interactions and behavior. Terms such as 'customer,' 'employee,' 'manager,' 'rude,' 'staff,' and 'worker' highlight customer experiences with employees, including their behavior and attitude. This topic reflects sentiments about staff professionalism and service quality.
- **Location and Food Quality**: The third topic gathers terms related to the physical location and quality of the food served. Words like 'fry,' 'chicken,' 'soda,' 'cold,' and 'sandwich' indicate issues about food products, while 'place,' 'table,' 'dirty,' 'bathroom,' refer to environment problems. This topic is concerned with the dining experience from the perspective of food quality and location-related aspects.

By categorizing the terms in this manner, the topics help in understanding specific areas of customer feedback, allowing for targeted improvements in service delivery, staff training, and overall dining experience.

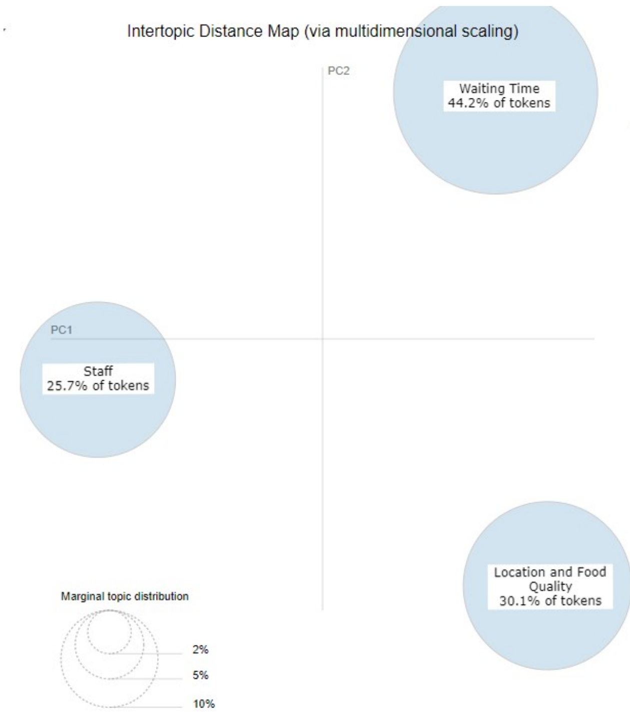


Figure 6. Intertopic Distance Map

Figure 6 represents the topics in a two-dimensional space. It reveals that the topics generated by the LDA model are well-separated, indicating clear distinctions between the identified themes. This separation allows for a straightforward interpretation of each topic's unique content and focus.

In particular, the 'Waiting Time' topic encompasses 44.2% of the tokens, making it the most prevalent topic. This indicates that waiting time is the most frequently mentioned issue in the reviews, highlighting its significant impact on customer experience. It is followed by the 'Location and Food Quality' topic, which accounts for 30% of the corpus, and the 'Staff' topic, which holds a 26% share.

The well-defined separation of topics in the visualization highlights the effectiveness of the model in distinguishing between different themes. This spatial arrangement aids in understanding the relative prominence and distinctiveness of each topic, facilitating a clearer interpretation of the underlying patterns within the data.

## 6. Conclusions

This work provides a comprehensive analysis of customer reviews related to McDonald's fast food restaurants, focusing on sentiment analysis and topic modeling to identify key areas of concern. The sentiment analysis, conducted using Recurrent Neural Networks (RNN) and Long Short-Term Memory (LSTM) models, revealed that both models struggled with accurately classifying neutral sentiments but performed better with positive and negative sentiments, with the LSTM showing superior performance. Topic modeling through Latent Dirichlet Allocation (LDA) identified three main themes in negative reviews: waiting time, staff behavior, and location and food quality. These insights can guide McDonald's in addressing specific customer issues, ultimately enhancing service quality and customer satisfaction.

## References

[1] Yoshua Bengio, Ian Goodfellow, and Aaron Courville. *Deep learning*, volume 1. MIT press Cambridge, MA, USA, 2017.

[2] David M Blei, Andrew Y Ng, and Michael I Jordan. Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan):993–1022, 2003.