Incorporating Topic Membership in Review Rating Prediction from Unstructured Data: A Gradient Boosting Approach.

Nan Yang, Nikolaos Korfiatis, Dimitris Zissis, Konstantina Spanaki

Incorporating Topic Membership in Rating Prediction from Unstructured Data: A Gradient Boosting Approach.

Nan Yang[a], Nikolaos Korfiatis[a*], Dimitris Zissis[a], Konstantina Spanaki[b]

[a]*Norwich Business School, University of East Anglia*
Norwich Research Park, NR4 7TJ,
United Kingdom

[b]*Audencia Business School,*
Nantes, France

[*]*Corresponding author: n.korfiatis@uea.ac.uk*

Incorporating Topic Membership in Rating Prediction from Unstructured Data: A Gradient Boosting Approach.

Abstract

Rating prediction is an essential aspect of business analytics due to the evaluation ability it provides to decision-makers regarding service performance. This study presents a dynamic model incorporating values from topic membership, an outcome variable from Latent Dirichlet Allocation, with sentiment analysis in an Extreme Gradient Boosting (XGBoost) model to assess rating prediction. The results show that, by incorporating features from simple unsupervised machine learning approaches (LDA-based), we can achieve an 86% prediction accuracy on objective rating values. At the same time, a combination of polarity and single-topic membership can yield a higher accuracy when compared with sentiment text detection tasks both at the document and sentence levels. Using Shapley Additive Values, we identify the additive predictability of topic membership values in conjunction with sentiment-based methods; based on a dataset from customer reviews about food delivery services. Incorporating performance features from verbatim text fields, especially in areas such as service quality measurements and customer satisfaction modelling, is a critical task in analytics that focuses on improving rating score predictions and demand forecasting.

Keywords: Latent Dirichlet Allocation, Sentiment Analysis, Machine Learning, Online Reviews, XGBoost

Statements and Declarations:
The authors disclose no financial or conflicting interests.

1 Introduction

The impact of customer feedback in the form of online reviews is an everlasting theme in the literature, and it has been found to have a significant impact on sales (Li et al., 2019; Z. Zhao et al., 2019) since customer reviews have been seen by consumers as embodiments of experience-specific information regarding the quality of products or services. The rating score from customer feedback represents customer overall satisfaction or dissatisfaction directly, thus affecting customers' purchase behaviour. As the antecedents of customer rating have been extensively researched, the rating prediction problem attracts more attention. Compared with using the characteristics of reviews (i.e., review length) for rating prediction, studies predicting rating scores using the textual contents within customer reviews are getting more popular. The textual content associated with the rating score is provided to justify the latter, thus explaining the underlying rationale, which offers consumers opportunities to express themselves freely rather than feel restricted by pre-defined subareas defined in the user interface (Büschken & Allenby, 2016).

The review rating score prediction problem is considered to originate from the sentiment classification task of classifying reviews as thumbs up (recommended) or thumbs down (not recommended) (Pang et al., 2002; Qiu et al., 2018). Several studies suggest a high level of consistency between a review's rating score and its textual justification (Hu et al., 2014; Qiu et al., 2018). Even the significant variation of ratings could be explained by customer sentiment statistically (Geetha, 2017), the numeric rating scores in customer reviews do not equally represent customer sentiment, as ratings cannot fully capture the polarity information appearing in reviews due to the limitations of the rating scale itself (Ghose & Ipeirotis, 2011).

Online reviews are widely adopted for companies to understand how customers perceive the quality of products or services. As demonstrated by Tirunillai and Tellis (2014), multidimensionality exists in quality measurement. Parasuraman Parasuraman, Zeithaml, and Berry (1988) developed an instrument called SERVQUAL to measure service quality using multiple dimensions, including reliability, tangibles, responsiveness, assurance and empathy. Similarly, there are latent dimensions within the textual content in online reviews (Tirunillai & Tellis, 2014). The predictive power of these dimensions has been examined by several studies. Nikolaos Korfiatis et al. (2019) illustrated that the extracted dimensions in the review text could add predictive accuracy to overall customer satisfaction. Xu (2020) also revealed that the textual factors in the review text have an asymmetric influence on customer satisfaction.

The latent dimensions can be discovered using topic models, which have been applied in various business areas to identify different themes from customers' textual feedback. Using topic models, we are able to discover the association and connection between a rating score and a review text, which could assist businesses in understanding the reasons driving different levels of customer satisfaction as well as the multidimensionality of the service's outcome (Korfiatis et al., 2019). In topic models, mix-membership models indicate that a document is considered a mixture of multiple topics, in which each word belongs to one single topic. Compared with single-membership, which allocates each document to only one topic, mix-membership models allow each document to cover multiple topics. We are able to extract the latent topics (dimensions) as well as how much each document is associated with each latent topic, which is the topic membership. LDA allows us to compute the topic membership, and that contained in unstructured data could increase the accuracy when predicting rating scores (Korfiatis, Stamolampros, Kourouthanassis, & Sagiadinos, 2019; Tirunillai & Tellis, 2014).

Therefore, in this study, we aim to evaluate the ability of topic membership to explain and predict overall customer satisfaction as well as explore how topic membership could be collaborated with customer sentiment to enhance the prediction accuracy. Specifically, using machine learning, we examine *how topic models can be used in conjunction with sentiment analysis to quantify customer feedback.* In addition, given the limitations and the issues with domain-specific sentiment analysis, we evaluate whether *topic modelling can provide a better alternative than sentiment analysis* in review rating prediction. The latter has significant implications, as it removes the necessity for employing domain-specific dictionaries and other approaches that are not adjusted to the particular vocabulary of the business domain.

To address these objectives, we perform a large-scale machine learning analysis on a dataset of 1,810,831 customer reviews from 12,153 restaurants on *JustEat*, a popular online food delivery platform in the United Kingdom. Our analysis is multi-faceted. We design and validated two experiments. First, a binary classification task is formed (using a rating score cut-off) to have a robust evaluation of the performance of topic membership as well as its comparison with customer sentiment (at both document-level and sentence-level). We also compare the predictive abilities across topic memberships of all topics. Second, the task of predicting the actual rating score is proposed. We combine each topic membership separately with the polarity score and compare the additive predictability of each topic in predicting the rating score. We perform a posthoc analysis for robustness by incorporating these two features as covariates using a gradient boosting model using XGBoost—an established machine learning technique. Using the corresponding Shapley additive explanation values (SHAP), we

identify the contribution to the prediction of the rating value by each topic/sentiment combination separately.

Our study contributes to the analytics literature by demonstrating how these two different approaches of incorporating features from unstructured data can be used in tandem to predict and explain customer satisfaction. Our findings can lead to faster and more accurate managerial insights for businesses since topic-based rating prediction can uncover multidimensional aspects of service quality that cannot be captured by sentiment analysis. Thus, it can explain customer rating scores and highlight service aspects within customer textual comments to facilitate businesses' understanding of customers' perceived quality towards products or services, which affects customers' purchase decisions (Yeo et al., 2022). Beyond customer reviews from review websites, there are also various sources of customer feedback used in business analytics cases, such as online forums and social media, which do not contain rating scores, or the rating score is incomplete for some of these dimensions. Our study can also extend to these sources.

To this end, the rest of this paper is organised as follows: Section 2 reviews the literature on sentiment analysis and topic modelling on customer reviews and how they are applied to rating prediction. Section 3 demonstrates our data sample, models, and metrics for model performance evaluation. Section 4 details the corpus pre-processing procedures and model parameter selection; then, it displays the results for these experiments. Section 5 summarises the analysis and discusses the implications of the results for researchers and practitioners. The study concludes in Section 6 with limitations and future research directions.

2 Literature review

2.1 Sentiment analysis

Sentiment analysis is a celebrated computational analysis method in business and management used for detecting customer attitudes and feelings expressed within an unstructured part of customer feedback through natural language processing (NLP) techniques. It is comprised of two main tasks: (a) polarity detection—the identification of whether the text is positive or negative and (b) affect detection—the feelings and emotions expressed in the written communication. Polarity detection is the most common approach applied in sentiment analysis, mainly thanks to the ease of labelling the large textual corpus concerning consumer interaction with company touchpoints (e.g., via social media). In that respect, the literature treats polarity as either an ordered categorical (with the text classified as positive/neutral/negative) or a

continuous numerical score of an asymmetric continuum ranging from a normalised negative algebraic value to a positive algebraic value (e.g., -1 to +1). Depending on the task at hand, sentiment can be calculated at the document or sentence levels. The latter also produces a confidence band that can produce more reliable prediction outcomes if sentiment is operationalised as input.

Apart from document-level and sentence-level sentiment analysis, aspect-level sentiment analysis is also discussed in the past literature. In many situations, it requires more investigation at the aspect level to identify entities and the associated aspects and sentiments (Do et al., 2019). For instance, companies would like to identify what aspects of their products attract or dissatisfy customers from customer reviews to improve products (Birjali et al., 2021). There could be two types of aspects: explicit aspects and implicit aspects. The former represents those aspects that are directly mentioned in the text; the latter illustrates those aspect terms that do not appear in the text but are implied by other terms (Hu & Liu, 2004). Several Machine learning approaches have been employed to extract the implicit aspects (Bagheri et al., 2013; Quan & Ren, 2014).

Various sentiment analysis techniques are discussed in the literature (Liu, 2010, 2012; Al-Natour & Turetken, 2020; Yadav & Vishwakarma, 2020). These can be summarised into two major types: lexicon-based and machine learning approaches. The lexicon-based sentiment approach uses a bag-of-words model that requires a dictionary consisting of predefined words or phrases assigned by negative or positive values. The other popular approach considers sentiment analysis as a pattern recognition problem and utilises machine learning techniques for classification tasks or predictions (Ghiassi & Lee, 2018). Marshan et al. (2020) proposed a TDS (Topic Document Sentiment) model, which is an unsupervised machine learning method based on the JST (Joint Sentiment Topic) model and LDA. They used the proposed model to discover the sentiment at the document, topic and word levels. Compared with sentiment terms (usually adjectives) from the lexicon-based approach, machine learning techniques can extract broader and more comprehensive features about several aspects of the text, including nouns and verbs expressing descriptions and attitudes towards these objects (Liu, 2012).

In addition, the hybrid approach, which combines lexicon-based approach and machine learning approach attracts more attention as it can integrate the advantages of machine learning approach (high accuracy and flexibility) and that of lexicon-based approach (stability) (Birjali, Kasri, & Beni-Hssane, 2021). For instance, Marshan, Kansouzidou, and Ioannou (2020) developed a hybrid model combing the lexicon-based and machine learning approaches to detect customer sentiment contained in reviews from an e-commerce platform. Three machine

learning approaches are selected, including the Naïve Bayes, KNN (k-nearest neighbours) and SVM (Support Vector Machine). Results showed that the Naïve Bayes had the best performance as a classifier. Elshakankery and Ahmed (2019) proposed a hybrid model HILATSA, representing Hybrid Incremental Learning approach for Arabic Tweets Sentiment Analysis to detect the sentiment in tweets. It is a semi-automatic learning system which will update the lexicon to keep it up to date.

Sentiment analysis, as a useful tool for detecting customers' emotions and attitudes, has been widely applied in user-generated content (e.g., online reviews), and several studies have examined the importance of customer sentiment in understanding the relationship with customer ratings, predicting sales, and identifying fraudulent reviews (Y. Zhao et al., 2019; Z. Zhao et al., 2019; Kumar et al., 2022). Innovative applications combining machine learning and bag-of-words-based approaches have been applied in practice. Dey et al. (2018) proposed a system that could generate Senti-N-Gram, an n-gram sentiment dictionary, and proposed an algorithm to extract the sentiment scores for n-grams from a random corpus consisting of review text as well as numerical ratings. This approach showed better performance than an existing unigram-based approach (VADER) and another n-gram-based approach (SO-CAL). Recent studies have also applied machine learning approaches to expand dictionary coverage. For instance, Sharma and Dutta (2021) proposed a framework called *SentiDraw*, which calculates the sentiment score for each word from customer reviews based on the rating distribution. Then it was combined with Support Vector Machine (SVM) to achieve better polarity determination.

2.2 Topic models on user-generated content

Online reviews, functioning as the *"voice of the consumer"*, are a form of electronic word-of-mouth (eWOM); these play a critical role in affecting customers' decision-making process, behavioural intention, and product sales performance (Li et al., 2019; Verma & Yadav, 2021). It has been considered an unignorable information source for both customers and sellers, especially the textual content within customer reviews, which includes the textual description of the first-hand usage experiences of previous customers (Guo et al., 2017). The growing popularity of online reviews provides customers with the opportunity to express themselves naturally with unstructured data.

Customers' opinions in online reviews are multidimensional and may reflect different aspects, such as product-specific features or service aspect related evaluations (Büschken &

Allenby, 2016; Kim et al., 2020; Mai & Le, 2021). Therefore, to ascertain these latent dimensions from customer reviews, the topic modelling approach is applied widely, as topic models could discover patterns reflecting latent topics within a document from unstructured customer reviews. It assumes that documents consist of a set of topics, and each topic covers a mixture of words (Alghamdi & Alfalqi, 2015). There are a variety of topic models, including Latent Semantic Analysis (LSA) (Landauer et al., 1998), PLSA (Hofmann, 2001), and LDA (Blei et al., 2003). There are many extensions of LDA, including the Correlated Topic Model (CTM) (Blei & Lafferty, 2007) and the Structural Topic Model (STM) (Roberts et al., 2014).

Researchers either adopt existing topic models or proposed new variants of topic models to discover the multidimensionality of customer reviews. The latent dimensions contained in customer reviews are critical since they serve as the foundation for how customers evaluate service, brands and firms, thus affecting new product development or brand positioning. Kwon et al. (2021) employed a topic modelling approach and sentiment analysis to online customer reviews for airlines in order to identify customers' needs. They extracted six dimensions using LDA and identified several words that contained positive and negative emotions respectively. Tirunillai and Tellis (2014) extended LDA and employed the variant of LDA to customers' reviews from five markets and 16 brands. They identified the latent dimensions and ascertained the valence, dynamics and heterogeneity, etc., for strategy analysis. Büschken and Allenby (2016) developed a new model (Sentence-constrained LDA model) based on LDA. They believe that people tend to change their topics across sentences instead of discussing two topics in one sentence. They applied it to two datasets consisting of customer feedback from both restaurant and hotel industries, illustrating the helpfulness of the topic modelling approach for unstructured data. N. Hu, Zhang, Gao, and Bose (2019) Hu employed STM to customer reviews from the hotel industry in order to understand customer dissatisfaction from their complaints. They identified the top 10 latent dimensions related to customer dissatisfaction and how dimensions change across hotel grades. Customers of high-grade hotels mainly complained about service issues, while of low-graded hotels are more dissatisfied with facility-related problems.

2.3 Rating prediction

Given the importance of understanding customer satisfaction, rating prediction is a vital task. Several studies have examined the characteristics of online reviews (e.g., review length) to understand customer ratings (Ghasemaghaei et al., 2018; Lai et al., 2021). The information

contained in textual comments also plays an important role in understanding customer ratings. Therefore, how to utilise review texts to predict customer rating scores has been a popular topic.

Based on whether the review text is provided in the prediction, the rating prediction task is mainly divided into two categories: (a) personalised rating prediction and (b) review-aware rating prediction. The first focuses on predicting users' rating scores over unrated items using their previous rating behaviours, which is widely explored in the recommendation system field (Zhang et al., 2016; Cheng et al., 2018). Latent factor models, including matrix factorisation, are applied widely and successfully for this type of rating prediction. Customers' textual comments could be corroborated to model user interests and item features (Tan et al., 2016; Zhang et al., 2016) and to improve the accuracy of rating prediction models. The second concentrates on understanding customers' rating scores by discovering valuable information from the provided review text. The direct relationship between sentiment and ratings has been confirmed in previous studies (Hu et al., 2014; Geetha et al., 2017). Table 1 summarises the two categories of rating prediction tasks, which approaches they adopt to extract features from the review text, and what prediction models they adopted.

Table 1 Overview of existing literature in rating prediction

Type	Task	Approach	Description	Prediction model	Indicative studies
Personalised rating prediction	Extracting contextual features and combining them to rating prediction	Topic model extensions	These studies were based on topic models, and combined topic modelling approaches with the latent factor models.	Self-proposed model	(McAuley & Leskovec, 2013; Zhang & Wang, 2016; Cheng et al., 2018)
	Extracting complex features and improving rating accuracy	Deep learning	These studies built deep learning architectures with multiple layers.	Self-proposed model	(Seo et al., 2017; Xing et al., 2019)
	Exploratory relationship with rating	Sentiment analysis	These studies discovered the direct impact of sentiment on review ratings and obtained the preliminary relationship between sentiment and ratings.	Fix effects model/linear regression	(Hu et al., 2014; Geetha et al., 2017)
	Rating prediction	Sentiment analysis	These studies applied sentiment analysis to review text and adopted the extracted sentiment features for rating prediction.	Ridge regression/linear regression	(Qu et al., 2010; Y. Zhao et al., 2019)
	Rating prediction	Topic models	This study applied LSA to discover positive and negative attributes from reviews and used these attributes as	Text regression	(Xu, 2020)

Review-aware rating prediction			independent variables in text regression to predict overall satisfaction.		
	Rating prediction	Both sentiment analysis and topic models	These studies applied LDA to exploit several dimensions from review text and combined them with sentiment detected by sentiment analysis to explain the overall satisfaction.	Linear regression/multinomial regression	(Xiang et al., 2017)
	Rating prediction	Self-proposed model	This study added constraints to the LDA model and proposed a sentence-constrained LDA model, and combined it with rating data.	Self-proposed model	(Büschken & Allenby, 2016)

Büschken and Allenby (2016) employed a variant of LDA to extract latent topics and to predict customer ratings using a dataset of customer reviews from Italian restaurants. It is examined that topic membership could be a meaningful device to explain customers' ratings scores. They used a latent cut-point model to examine the relationships between customer satisfaction and the topic membership of 8 topics. Xu (2020) employed LSI in the content of customer reviews from the hotel industry and identified 8 positive and 17 negative factors. Text regression was conducted using the vector space of each textual review to examine how it can affect overall customer satisfaction. The asymmetric effects were found from the results representing that not all positive textual factors affected customer satisfaction positively. Nikolaos Korfiatis et al. (2019) adopted STM to online reviews from airline passengers and extracted latent dimensions of service quality from the textual content. Together with the predefined subcategories by the online platforms, the latent dimensions could add the ability to predict overall customer satisfaction. These studies provided us with another way to predict customer ratings in addition to using the predefined subscales by companies or platforms and prove the ability of topic membership to predict and explain customer ratings.

3 Data and methods

For this study, we follow the Cross-Industry Standard Process for Data Mining (CRISP-DM), proposed by Wirth and Hipp (2000), aiming to convert business problems into data mining projects that could be carried out and applied regardless of the type of technologies and industries. We illustrate "Business Understanding" and "Data Understanding" in Section 1 and Section 2. We move to "Data Preparation" in this section and "Modelling" in Section 4.

3.1 Dataset

Our data considers textual reviews of UK customers and is collected from *JustEat*, the most popular online food delivery service provider in the UK, with more than 68% of the market share of online orders[1]. Besides, *JustEat* could provide customers with each review, including rating score and review text from previous customers who purchased in the specific restaurant, while other competitors (e.g., *Deliveroo* and *UberEats*) only display the overall rating score and the number of reviews of the specific restaurant. After customers order and receive the delivered food, they are invited to leave customer reviews describing the entire experience with the food delivery. Potential customers searching for a restaurant can find these reviews on restaurants' pages, which can be of assistance to them. We collected customer reviews written

[1] Statista Global Consumer Survey – Brand Report, 2021

in English and published them on their website from January 2016 to November 2021. Generally, each review should have a numerical rating score with textual justification. *JustEat* adopts a 6-point scale rating system in which customers give a rating from 1 to 6 stars for three service categories: food quality, delivery time, and restaurant service. The final rating score shown to other consumers when reading these reviews is calculated as the (simple) average of the individual ratings of these three categories.

The textual justification provided by the customer considers all three service categories; therefore, the average rating provides intervals between the minimum and maximum rating scores. To make our analysis more meaningful, we select reviews with textual comments from customers and filter out those that only contain rating scores. Additionally, each review length is constrained terms of length between 15 and 200 words[2]. In total, our sample contains 1,810,831 customer reviews from 12,153 restaurants. Using the density distribution of the ratings provided in the dataset, we used the median of the rating scale (3.5 stars) as the boundary (blue line Fig.1) for separating the positive and negative reviews, given its even distribution in both classes.

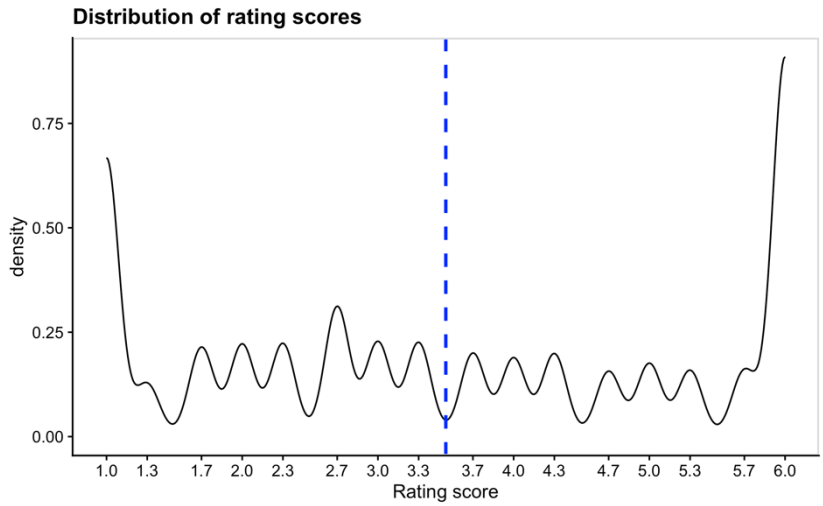

Fig. 1 Distribution of rating scores within the reviews in our sample. The blue dashed line outlines the median rating score.

Fig. 1 shows the distribution of the rating scores for our sample. This indicates that the percentages of extreme ratings are significantly higher than others. The average rating score is 3.57 ($sd = 1.82$), which is slightly positive. As to the actual distribution of the scores, the highest percentage occurs in 6-stars ratings, amounting to 20.9%, followed by a 1-star rating, which has the second-highest proportion (15.3%). In total, the proportions of positive (49.19%)

[2] A winsorization procedure was followed for the maximum values considering that any reviews about 200 words were above the 95% quantile of the distribution of review word length.

and negative reviews (50.81%) in our sample are similar, which means that our sample is balanced.

3.2 Sentiment analysis

Sentiment analysis is generally measured through polarity, which measures the positive/negative intent expressed in the text. It is generally calculated by various techniques related to the bag-of-words approach; however, other studies in the literature have also applied more complicated models. For polarity calculation, the standard method is to use a lexicon-based approach with domain-specific or general dictionaries (or lexicons). These lexicons can be compiled manually or acquired automatically. The function we adopt to calculate polarity is based on subjectivity lexicons, which contain a list of terms connected with particular emotional states. For instance, the word 'awful' relates to a negative state, while the word 'excellent' is associated with a positive state.

We employ the subjectivity lexicon from Hu and Liu (2004) to calculate polarity, including approximately 6,800 prior-labelled words, which have been identified by benchmarking these terms on a large dataset of online consumer reviews. For identifying polarity in the text, a cluster of terms (x_i^T) that contains four words before and two words after the polarised word has been used to introduce context. Words in the cluster that do not have value are called neutral words and are tagged as x_i^0, which affect word count (n). In addition, there are words that do not have emotion but have an influence on the emotional context, such as valence shifters. Amplifiers (x_i^a)/De-amplifiers (x_i^d)/ are words which can increase/decrease the emotional intent of words, which are given a weight for calculation (Rinker, 2020). The context is defined as follows:

$$x_i^T = \sum((1 + c(x_i^A - x_i^D)) * w(-1) \sum x_i^N,$$

Where:

$$x_i^A = \sum\left(w_{neg} * x_i^a\right), x_i^D = \max\left(x_i^{D'}, -1\right),$$

$$x_i^{D'} = \sum\left(-w_{neg} * x_i^a + x_i^d\right),$$

$$w_{neg} = \left(\sum x_i^N\right) mod\ 2.$$

The polarity score is calculated as:

$$\delta = \frac{x_i^T}{\sqrt{n}}$$

We constrain the polarity score at (-1,1) using a transformation formula: $\left[\left(1-\frac{1}{\exp(\delta)}\right)*2\right]-1$. We use consumers' original review comments before Part-of-Speech tagging and stop words removal because the absence of words will decrease the accuracy by affecting the density of keywords.

3.3 Latent Dirichlet Allocation (LDA)

The LDA model proposed by Blei et al. (2003) is an unsupervised learning model based on Bayesian inference. Its underlying principle is exchangeability. Compared with latent semantic indexing (LSI) and probabilistic LSI (pLSI) models, LDA considers the exchangeability of both documents and words. LSI applies statistical computations to a large corpus of text to extract and represent the contextual-usage meaning of words (Batra & Bawa, 2010). LSI adapts Singular value decomposition (SVD) into the term-document matrix to achieve dimensionality reduction (Zelikovitz & Hirsh, 2001). Deerwester, Dumais, Furnas, Landauer, and Harshman (1990) demonstrated that several basic linguistic notions (e.g., synonymy and polysemy) could be captured by the linear combinations of the tf-idf features, which are derived by LSI. However, the biggest weakness of LSI is the lack of a satisfactory statistical foundation. Subsequently, the probabilistic LSI (PLSI), with a solid statistical foundation using a probabilistic method replacing SVD, is proposed by Hofmann (2001) to address the weakness of LSI. It considers each word as a sample from one mixture model, in which the mixture components (multinomial random variables) are considered "topics". In pLSI, a list of mixing proportions for these mixture components is used to represent each document. However, no probabilistic model at the document-level is provided, meaning the numbers from each list are not from any generative probabilistic model, leading to overfitting problems seriously (Blei et al., 2003).

LDA is a generative probabilistic model that can deal with sparse vectors of discrete data, including bag-of-words in text data and image features. For text data, the core assumption is that each document is considered a random mixture of latent topics, while a multinominal distribution over words represents each topic. LDA is based on the assumption that the author of each document would have the same probability of using exact words when writing the same "*topic*". LDA is a generative probabilistic model, simulating an author's process of producing a document. In this process, the probability of writing a word is related to the topic that is written about. However, if two authors have different words to write for the same document, the distribution of words might change to an unrelated topic by making an inaccurate inference.

Every document is created by a list of hypothetical and unobservable *'topics'*. Each document is assumed to be presented by a mixture of topics that reflect distributions sharing common Dirichlet priors. In a single document, the probability of each topic is between 0 and 1, with the sum of them amounting to 1. The extent to which documents are associated with topics is considered document-topic proportions, also known as topic membership. The latent topics are considered as the distributions over a fixed vocabulary in which each word has a possibility of belonging to each latent topic.

In this study, each review is a single document. We index each review as $r \in (1,2 \dots, R)$. K presents the number of topics, which is the primary input variable. The generation process is summarised as follows:

For each topic $k \in (1,2 \dots, K)$, draw a Dirichlet distribution over the vocabulary V, $\beta_k \sim Dir(\eta)$.

For each review, r, choose $\theta_r \sim Dir(\alpha)$.

i. For each word w_a, from review r, a topic assignment is drawn from a multinominal distribution over $\theta_r, z_{r,a} \sim Mult(\theta_r)$, where $z_{r,a}$ represents the word-specific topic assignment.

ii. The observed word, $w_{r,a}$ is drawn from $Mult(\beta_{z,r,a})$, where $w_{r,a} \in (1,2 \dots V)$.

The joint distribution of all unobserved variables and observed variables is expressed as follows:

$$p(\beta_K, \theta_R, z_R, w_R|\alpha, \eta) = \prod_{k=1}^{K} p(\beta_K|\eta) \prod_{r=1}^{R} p(\theta_r|\alpha) \prod_{a=1}^{N} p\left(z_{r,a}|\theta_r\right) p\left(w_{r,a}|z_{r,a}, \beta_{r,k}\right)$$

Fig. 2 depicts the graphical model of the LDA process in plate notation. The shaded nodes present the only observed variables $w_{r,a}$, which represents the ath word in review r. A is the total number of words in each review. Each review could be represented as a mixture of topics. θ denotes the review-topic distribution, which indicates how much each review is related to topics. β denotes the per-review topic–word distributions. Sparsity problem caused by the large vocabulary occurs in many document corpora. The "smooth" method is usually adopted to avoid assigning zero probability to words that are from new documents but do not appear in documents from the training corpus ((Blei et al., 2003)). Instead of the commonly used method-Laplace smoothing, LDA places the Dirichlet priors on the multinomial parameters. Therefore, α and η, as two Dirichlet parameters, denote the smoothing of topics with reviews and words within topics, respectively (Syed & Spruit, 2017).

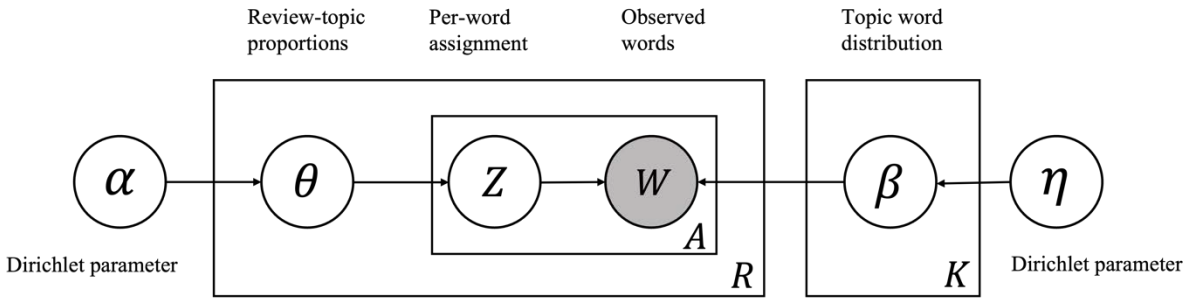

Fig. 1 LDA model process using plate notation (Adopted by Blei et al., 2003)

The topic-word distributions and the coefficients for documents cannot be observed and are estimated by a learning algorithm, such as expectation propagation (a higher-order variational algorithm) and the Markov chain Monte Carlo algorithm (Griffiths & Steyvers, 2002; Minka & Lafferty, 2002; Porteous et al., 2008). For model inference and parameter estimation, we adopted Gibb's sampling to compute the approximations to the posterior distribution of the hidden variables in the model, which is the core inferential problem in LDA. Compared with the convexity-based variational approach introduced by Blei et al. (2003), Gibb's sampling could achieve higher accuracy by approaching the asymptotically correct distribution (Porteous et al., 2008).

3.4 Extreme Gradient Boosting (XGBoost)

To examine the ability of two approaches to predict rating scores, we conduct a set of two-stage experiments: binary classification and rating prediction. Classification is used to identify which category the new observation belongs to, while prediction involves making future estimations based on current data behaviour patterns (Brintrup, 2021). Extreme Gradient Boosting (XGBoost) is a highly effective scalable tree-boosting system. It has achieved state-of-the-art results on a wide range of machine learning challenges because of its effectiveness, flexibility, and portability (Chen & Guestrin, 2016).

Giannakas et al. (2021) compared the ability of XGBoost with a 4-hidden-layers Deep Neural Network (DNN) when making a prediction of the team performance. The results revealed that both the learning accuracy and prediction accuracy of XGBoost are higher than DNN. Wu et al. (2021) applied five different datasets to examine the performances of XGboost and Multiple-layer Perceptron Neural Network for binary classification tasks. The results demonstrated that XGBoost performed generally better than the neural network and significantly better when the overlapped samples increased. Khanam et al. (2021) evaluated the performance 7 algorithms, including Logistic Regression, XGBoost, KNN, Naïve Bayes, Decision, SVM and Random Forests, when performing classification tasks for fake news

detection. It is examined that XGBoost depicted the highest accuracy than other algorithms. Rao et al. (2021) also compared the performance of several algorithms as classifiers, including XGBoost, Logistic Regression, Random Forest, Decision Tree, Multinomial Naïve Bayes and Bernoulli Naïve Bayes to perform the binary classification task of detecting fake news. They demonstrated that XGBoost could simultaneously provide an excellent mix of prediction and processing speed. After fine-tuning hyperparameters, it could achieve the highest accuracy than other methods. Apart from classification, another study by Yan et al. (2022) examined the power of XGBoost to make predictions in the health field. They compared XGBoost with a multivariate logistic regression model and found that the former performed better predicting the risk of death with one specific disease. Due to the better performance of XGBoost compared with other algorithms, we believe that XGBoost is an appropriate approach for classification and prediction tasks.

The gradient boosting approach is described as follows: Assume a dataset $D = \{(x_i, y_i): i = 1, \dots, n, x_i \in R^m, y_i \in R^n\}$, with n instances and m features.

$$\widehat{y_i} = \sum_{k=1}^{K} f_k(x_i), f_k \in F$$

where K represents the total number of trees, $f_k(x_i)$ is the predicted value of i-th sample in the k-th tree, F is the function space consisting of all CARTs (regression or classification trees in XGBoost), and $\widehat{y_i}$ is the predicted value, for instance, i.

The set of functions $f(k)$ could be learned by minimising the objective function, which consists of training loss $L(\theta)$ and regularisation term $\Omega(\theta)$.

$$O(\theta) = L(\theta) + \Omega(\theta)$$

$$= \sum_{i}^{n} l(y_i, \widehat{y_i}) + \sum_{k=1}^{K} \Omega(f_k)$$

Where l is the training losing function, measuring the difference between the predicted value $\widehat{y_i}$ and the observed value y_i The regularisation term Ω could control the model complexity to avoid overfitting.

The tree model could be trained using an additive strategy.

$$\widehat{y_i}^{t} = \widehat{y_i}^{t-1} + f_t(x_i)$$

Therefore, the objective function at step t is changed as follows:

$$O^{(t)} = \sum_{i=1}^{n} l\left(y_i, \widehat{y_i}^{t-1} + f_t(x_i)\right) + \Omega(f_t) + constant$$

By using the second-order Taylor expansion, we simplify the equation as below:

$$O^{(t)} = \sum_{i=1}^{n} [\left(l\left(y_i, \hat{y_i}^{t-1}\right) + g_i f_t(x_i) + \frac{1}{2} h_i {f_t}^2(x_i) \right] + \Omega(f_t) + constant$$

Where g_i and h_i are represented as follows:

$$g_i = \partial_{\hat{y_i}^{t-1}} l\left(y_i, \hat{y_i}^{t-1}\right)$$

$$h_i = \partial^2{}_{\hat{y_i}^{t-1}} l\left(y_i, \hat{y_i}^{t-1}\right)$$

In XGBoost, the complexity is defined as:

$$\Omega(f_k) = \gamma T + \frac{1}{2}\lambda ||w||^2$$

Where γ and λ are regularisation parameters, T represents the number of leaves and w are scores on leaves. By defining $G_j = \sum_{i \in I_j} g_i$ and $H_j = \sum_{i \in I_j} h_i$ and expanding the regularisation term, the objective function is re-formulated as:

$$O^{(t)} = \sum_{j=1}^{T} [G_j w_j + \frac{1}{2}(H_j + \lambda) w_j^2] + \gamma T$$

The best w_j^* and the best corresponding value could be computed as

$$w_j^* = -\frac{G_j}{H_j + \lambda}$$

$$O_j^* = -\frac{1}{2} \sum_{j=1}^{T} \frac{G_j^2}{H_j + \lambda} + \lambda T$$

Given that enumerating all possible trees is not intractable, the tree is optimised on one level at a time by splitting leaves and producing a gain score:

$$Gain = \frac{1}{2}\left[\frac{G_L^2}{H_L + \lambda} + \frac{G_R^2}{H_R + \lambda} - \frac{(G_L + G_R)^2}{H_L + H_R + \lambda} \right] - \gamma$$

3.5 Model performance measurement metrics

In our study, we use two types of metrics to compare the performance of the classification and prediction models. For the classification, a standard approach utilising a confusion matrix is used to represent the dispositions of the test dataset in a 2 x 2 setting (true positive, true negative, false negative, false positive).

The true positive rate, also known as recall or sensitivity, is calculated as:

$$Sensitivity = \frac{TP}{TP + FN}$$

The false negative rate, also known as specificity, is estimated as:

$$Specificity = \frac{TN}{TN + FP}$$

Precision is calculated as:

$$Precision = \frac{TP}{TP + FP}$$

An equal weighted combination of these two metrics can be reflected on the F1-score, which can be calculated as:

$$F1\ score = 2 * \frac{Precision * Recall}{Precision + Recall}$$

The false positive rate (FPR) is equivalent to $1 - Specificity$. The ROC curve is a two-dimensional plot that shows sensitivity on the y-axis and $1 - Specificity$ (FPR) on the x-axis. The perfect one is located at point (0,1). The ROC curve starts from point (0,0) and ends at point (1,1). AUC is a method to compare classifiers and is calculated as the area under the curve, which is between 0 and 1. The model with a higher AUC performs better in classification than others, as it is known that *"the AUC of a classifier is equivalent to the probability that the classifier will rank a randomly chosen positive instance higher than a randomly chosen negative instance"* (Fawcett, 2006, p. 868).

In addition to the binary classification, we would perform regression using XGBoost. Thus, to determine the effectiveness of our model, two metrics will be calculated: Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE). MAE represents the average of the difference between the predicted value and the original value, which can be calculated by the formula below. A smaller MAE indicates a better model.

$$Mean\ Absolute\ Error = \frac{1}{N}\sum_{i=1}^{N} |y_i - \widehat{y_i}|$$

RMSE is popularly used in the literature, which represents the square root of mean squared error, as the equation below shows. It is similar to standard deviation and could measure how much residuals are spread out.

$$Root\ Mean\ Squared\ Error = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \widehat{y_i})^2}$$

All metrics are estimated and used in evaluating the performance of each model in the analysis.

3.6 Feature selection

Unlike simple models (e.g., linear regression), more complex predictive models (e.g., deep learning and tree-based models) are complicated to interpret. Shapley Additive explanations (SHAP) values, proposed by Lundberg and Lee (2017), could better interpret black-box models by computing Shapley values from coalitional game theory. The Shapley value is an explanation method based on solid theory, in which four axioms (efficiency, symmetry, dummy, and additivity) provide a reasonable foundation. It represents how much a feature contributes to the prediction for each instance compared with the average prediction of the trained model (Molnar, 2020). Inspired by cooperative game theory, the Shapley value is a method of fairly distributing pay-outs to players according to their contribution. In this study's circumstances, the 'players' represent the feature values, and the 'game' signifies the prediction task.

SHAP could explain the Shapley values as a linear model specified as:

$$g(z) = \emptyset_0 + \sum_{j=1}^{M} \emptyset_j z_j$$

where M is the maximum number of simplified input features. $z \in \{0,1\}^M$. When calculating the Shapley value, the value of z represents the status of the presence (used or not used) of the corresponding feature in prediction. $\emptyset_j$ is the Shapley value (the attribution of feature j). The Shapley value of feature j can be calculated as follows:

$$\emptyset(j) = \sum_{s \subseteq \{1,\ldots,M\} \setminus \{j\}} \frac{|S|!\,(M - |S| - 1)!}{M!} (v(S \cup \{j\}) - v(S))$$

Where S represents one subset of the simplified features included in the model. $v(S)$ is the total value for S. The marginal contribution of feature j is calculated as $v(S \cup \{j\}) - v(S)$.

SHAP could be a powerful method to interpret results from tree-based machine learning models (e.g., random forest and gradient boosted trees). In this study, SHAP demonstrates the feature importance by examining its marginal contribution to the model output, which provides local explanation and consistency globally.

4 Results

4.1 Baseline sentiment calculation

For each review, we calculate the sentiment polarity score (document-level and sentence-level polarity) using the original textual comment at the review level and provide a decimal between

-1 and 1. Fig. provides the distribution of document-level polarity scores in our sample. The biggest peak of the curve is in the middle. The average polarity score is 0.05, and the standard deviation equals 0.28. The most frequently occurring polarity values are clustered near the middle. The extreme polarity scores (close to 1.0 and -1.0) occur the least frequently, which is quite different from the distribution of customer rating scores (1), as extreme rating scores show the most frequent occurrence.

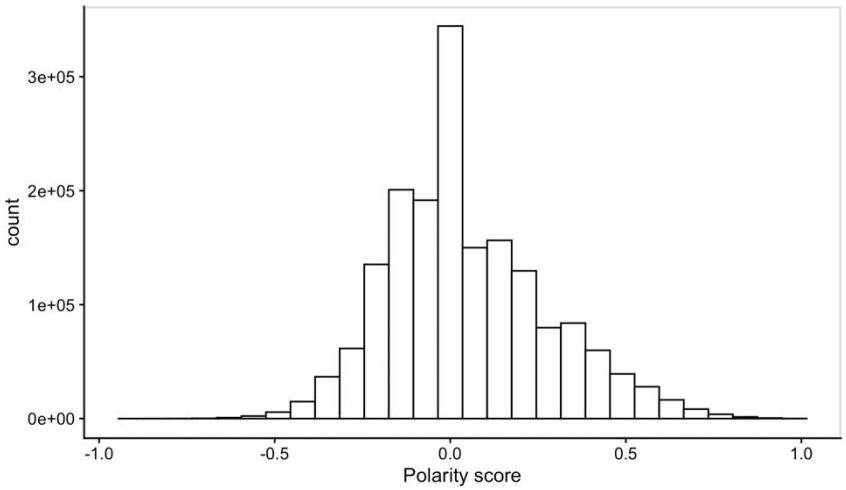

Fig. 2 Polarity score distribution for our sample

4.2 Corpus pre-processing for topic modelling

The textual content from customer reviews is pre-processed by following the standard procedural remedies, including (a) word tokenization (breaking sentences into a set of tokens), (b) exclusion of numbers and punctuations, (c) elimination of stop words, which includes both language stop words removal using the SMART stop word list and context-specific stop words exclusion, such as food vocabulary and restaurant brand names, and (d) selecting only adjectives, nouns, and adverbs from remaining words, since these words contain relevant information about products and product quality (Tirunillai & Tellis, 2014). This step is implemented by utilising part-of-speech (POS) tagging to keep the parts of speech that are meaningful as well as lemmatization, which derives the base forms of the words. (e) For low-frequency words (frequency of occurrence is less than 2% of the total amount of reviews), a pruning procedure is followed, reducing the number of reviews to *1,700,131*. The low-frequency words that convey highly specific semantic information are considered weak features in the corpus (Leeman, 2007). We followed the procedure of removing low-frequency words based on the study of Griffiths and Steyvers (2004).

4.3 LDA model estimation and hyperparameter tuning

After transforming textual data into a document-term matrix, we estimate the topic models and use a heuristic approach to evaluate the hyperparameter values that provide an ideal solution using the current parameter set. As shown in Fig. , there are two hyperparameters, α and η, which are two smoothing parameters controlling the sparseness of Dirichlet distribution. These two values and the number of topics K are required to be inputted for the LDA process. To determine the best number of K, many researchers have adopted trial and error procedures. A set of models was estimated with various values of K and the model producing the most meaningful topics was selected (Blei, 2012; Bastani et al., 2019). Based on the intrinsic nature of reviews from the OFD platform, we could infer that reviews are homogeneous and concentrated on only a few themes (e.g., food quality, delivery speed, and driver's attitude). As *JustEat* adopted a 6-star rating system, we estimate 13 LDA models with different values of K starting from 6 to 18.

Several researchers seek to find the best number by calculating the *'perplexity'* of the held-out test set, which is one intrinsic evaluation metric for language model evaluation. Perplexity algebraically equals the inverse of the geometric mean per-word likelihood (Blei et al., 2003).The lower perplexity indicates that the model predicts better for new test samples. However, there is a distinct drawback to using perplexity to evaluate the quality of the LDA model. The perplexity decreases as the number of topics increases (Koltcov et al., 2014). Chang et al. (2009) illustrated that producing ever finer partitions as the number of topics grew could make the model less helpful and reduce topic interpretability.

Therefore, to find the best number of K, two metrics are calculated, proposed by Cao et al. (2009) and Deveaud et al. (2014), to compare 13 LDA models. Cao et al. (2009) found that the best K is not only correlated with the size of the dataset but is also influenced by the inherent correlations within the corpus. They considered each topic as a semantic cluster, in which the similarity of each word is as small as possible, while the similarities among topics are expected to be large. Similar to the idea of clustering based on density, they aim to achieve a large similarity within the topic for more explicit semantic meaning while a small similarity among topics shows a stable topic structure. The procedures are as follows: First, the initial LDA model is estimated given an arbitrary K value. Second, they calculate the average cosine distance of the model, the model's cardinality, and all topics' density. Third, based on the cardinality, they re-estimate the LDA model and initialise sufficient statistics. If the direction of convergence is negative, topics with high densities will be applied as reference samples.

Otherwise, the seeded method will be adopted to initialise it. Then, repeat the second and third steps until the model's average cosine distance and cardinality converge.

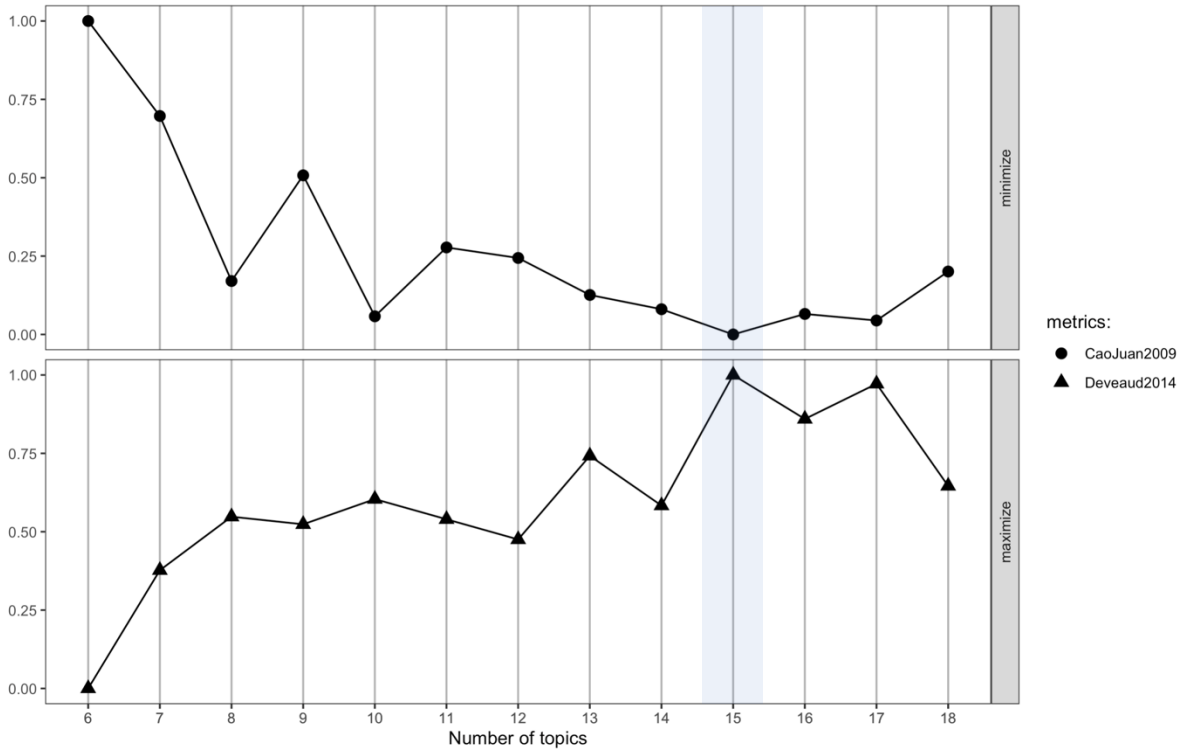

Fig. 1 3 Selection of the number of topics (K) for identifying the topic solution. Optimal K is identified by the shaded area.

In addition, Deveaud et al. (2014) proposed a simple heuristic approach to find the best number of topics when the information diverges between all pairs within the LDA model. Rather than the non-symmetric measure (Kullback-Leibler divergence), the symmetrised version of the Jensen-Shannon divergence is applied. Fig. depicts the performance of different LDA models using these types of metrics for the values of K (x-axis). The model achieves the best performance when the upper metric has the minimum value, or the lower one is maximised. Therefore, we select 15 as the optimal value of K.

4.4 Topic identification

Table 2 provides the K=15 topic solution for the review corpus, which is the optimised solution after topic number selection, as previously discussed. It represents 15 topics' top 7 loading words separately produced using the standard topic word probability (β) from the LDA estimation process. Several topics show positive intention, such as Topics #1, #2, #3, and #5, while some topics are more negative (i.e., Topics #11, #12, and #13). For instance, Topic #12 mainly talks about the delivery service from drivers about locating their addresses, and topic

#13 concentrates on late deliveries and long waiting times. Topic #3 focuses on customers' praise and subjectively positive descriptions of their takeaways.

Table 2 15 topics and their top 7 loading words in the topic solution.

Topic	Top 7 loading words
Topic 1	food cold stone longer late delivery driver
Topic 2	hot nice food lovely fresh tasty again
Topic 3	best place amazing delicious guy takeaway excellent
Topic 4	service back customer phone poor bad problem
Topic 5	food great always early delivery fast home
Topic 6	good quality portion large small price worth
Topic 7	meal only happy extra box thing instead
Topic 8	order wrong right issue correct number store
Topic 9	not more disappointed taste lot flavour same
Topic 10	food again warm free once hungry barely
Topic 11	never again money ever dry soggy hard
Topic 12	delivery driver where door address house man
Topic 13	late hour minute min half way later
Topic 14	time first last long few second next
Topic15	drink item missing bag refund order full

Add a paragraph of the topic membership

4.5 Incorporating topic membership in sentiment text detection

The literature review shows that sentiment and topic membership could be considered two devices to explain and predict customer satisfaction. We would like to examine how sentiment and topic membership can predict customer satisfaction empirically. As mentioned in Section 3.1, we consider the mid-point of the rating scale (3.5 stars) as the boundary to separate negative and positive reviews. Based on its rating score, each review in our sample is classified into two classes: *positive* and *negative*. Therefore, by classifying customer reviews into positive and negative, the question is transformed into a binary classification task. More specially, we would also examine the different abilities of document-level sentiment and sentence-level sentiment in this classification task. Therefore, we constructed three models (Table 3) with all target variables being rating (positive/negative) for the classification task. Model A and Model B include each review's calculated polarity score (document-level and sentence-level separately) as independent variables to predict the class. Model C adopted the topic membership of 15 topics from the topic solution of the LDA process as the predictors. The in-sample validation split was 80% for training and 20% for testing. For both classification tasks,

the XGBoost algorithm is followed. For hyperparameter selection, we limit the maximum depth of the tree to 2 and the maximum rounds of boosting iterations to 100 to obtain the optimal outcome.

Table 3 Three models' construction for the classification task

Target Variable		Polarity (Document level)	Polarity (Sentence level)	Topic membership (15 topics)	AUC
Rating (positive/ negative)	**Model A**	*			0.827
	Model B		*		0.838
	Model C			*	0.860

AUC scores are calculated using k-folding, and ROC curves are graphically presented in Fig. 5. As presented in Fig. 5, the dashed curve represents the ROC curves of Model A and Model B (using document-level polarity and sentence-level polarity separately), whose AUC scores are 0.827 (95% CI: 0.825–0.828) and 0.838 (95% CI: 0.837–0.840), respectively showing relatively good classification, as they are both larger than 0.8. The sentence-level polarity score has better performance than the document-level polarity score in classifying the two rating classes. The solid line is the ROC curve of Model C using topic membership (15 topics), demonstrating better performance than using polarity; the curve close to the upper left corner indicates higher accuracy. The AUC score is 0.860 (95% CI: 0.8587–0.8611), which represents a better ability to separate positive and negative classes. Only comparing the two models with polarity, solely using topic membership on its own, could increase the predictive accuracy.

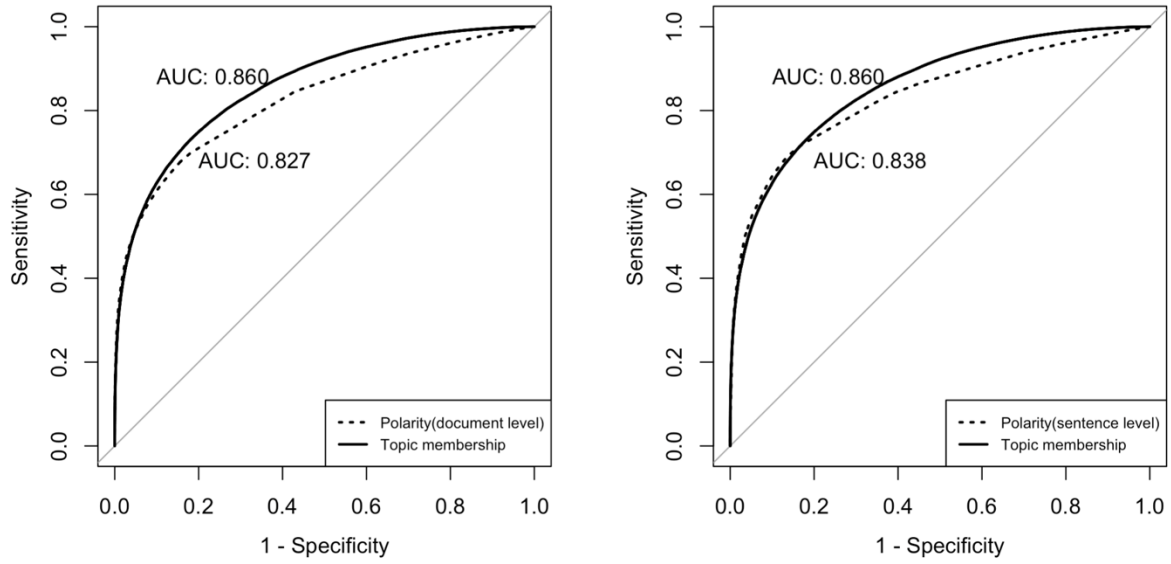

Fig. 5 AUC comparison (left-hand side) between Model A and Model C as well as AUC comparison (right-hand side) between Model B and Model C

As mentioned, topic membership, including all topics, performs better than sentiment in helping to classify the positive and negative reviews. These 15 latent topics extracted through the LDA process indicate various dimensions customers pay attention to contained in customer reviews. Some might be emotional and highly related to customers' ratings, while others might be more realistic. To examine the individual contribution of each topic to the predictive accuracy of binary classification, we construct and perform 15 models (with the same hyperparameters as models in the previous experiment) to classify positive and negative classes and include the topic membership for each topic as predictors. Demonstrates the AUC scores for the 15 models. Several models (Models #4, #7, #9, #10, #12, and #14) have relatively low AUC scores (close to 0.5) and do not help much in predicting customer attitudes. The remaining models have higher predictive performance for classification in some way, with higher AUC scores than 0.6. Among them, Model 3 has the highest AUC score (0.706) and presents the best performance in the binary classification task. The dimension that Topic #3 mainly talks about indicates the strongest relationship with customers' attitudes (positive or negative).

Table 4 AUC Comparison among 15 topics' memberships in the classification task

Predictor	Topic1	Topic2	Topic3	Topic 4	Topic 5	Topic 6	Topic7	Topic8	Topic9	Topic10	Topic11	Topic12	Topic13	Topic14	Topic15
Model 1	0.621														
Model 2		0.677													
Model 3			**0.706**												
Model 4				0.571											
Model 5					0.637										
Model 6						0.612									
Model 7							0.546								
Model 8								0.605							
Model 9									0.548						
Model 10										0.539					
Model 11											0.618				
Model 12												0.542			
Model 13													0.649		
Model 14														0.551	
Model 15															0.626

We already examined the model with topic membership and the model with sentiment only separately and proved that topic membership with all topics included could perform better than polarity only in classifying positive and negative classes. However, the combined predictive power of sentiment and topic membership has not yet been examined. As Topic #3 contributes most to the predictive accuracy of classification, we construct Model D and Model E with the integration of Topic #3 and polarity (two levels) as predictors for the classification. The former includes document-level polarity score and topic membership (only Topic #3), while the latter includes sentence-level polarity score and topic membership (only Topic #3) as predictors.

Models were performed with the same hyperparameters, and their ROC curves and AUC scores are displayed and compared with Model A and Model B, as shown in Fig. 6. The comparison (the left-hand side) of Model A and Model D showed that Topic #3 added as a new predictor together with document-level polarity could increase the AUC score to 0.842 (95% CI: 0.841–0.843). The comparison (the right-hand side) of Model B and Model E revealed that Topic #3, together with the sentence-level polarity, could increase the AUC score to 0.852 (95% CI: 0.851–0.854). It demonstrates that topic membership possesses the additional power to help sentiment predict customer attitudes.

Table 5 Two models construction with a combination of sentiment and topic (#3) membership

Target Variable		Topic membership (only Topic 3)	Polarity (Document level)	Polarity (Sentence level)	AUC
Rating (positive/ negative)	**Model D**	*	*		0.842
	Model E	*		*	0.852

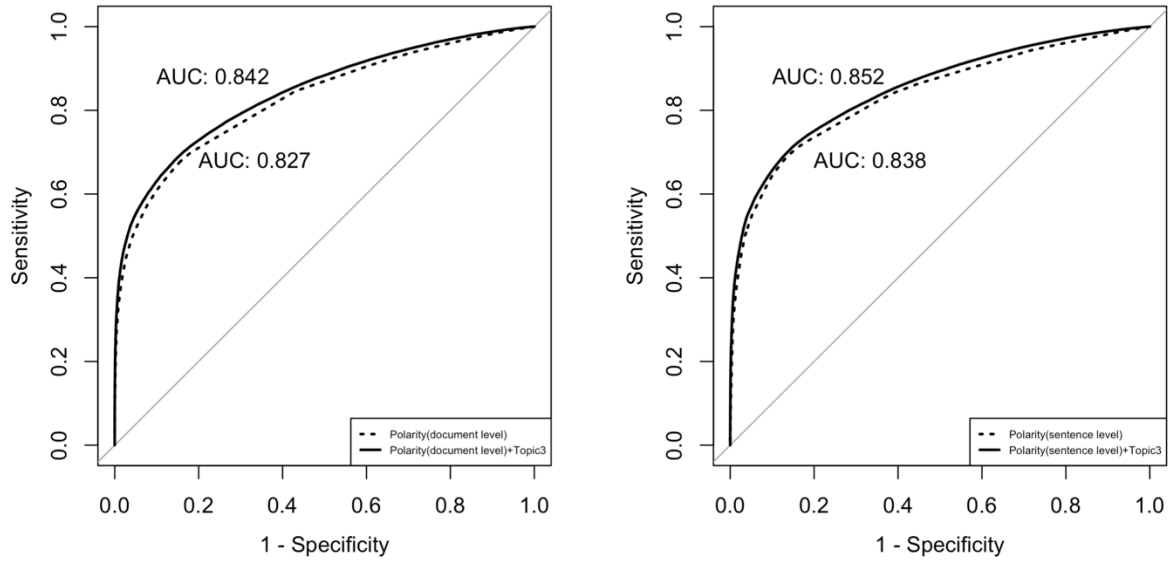

Fig. 6 AUC comparison (left-hand side) between Model A and Model D as well as AUC comparison (right-hand side) between Model B and Model E

The parameters that decide the model architecture are hyperparameters. We could find the ideal hyperparameters and improve our predictive accuracy through hyperparameter tuning. After tuning hyperparameters, we find a relatively better list of hyperparameters with a learning rate of 0.3, gamma of 0.2, maximum depth of a tree as 7, and the minimum sum of instance weight as 5 in a child, and the subsample ratio as 0.8. Finally, as Fig. 7 shows, we improve the AUC scores of models with polarity (both document-level and sentence-level) and Topic #3 as predictors to 0.845 (95% CI: 0.843–0.846) and 0.856 (95% CI: 0.854–0.857), respectively.

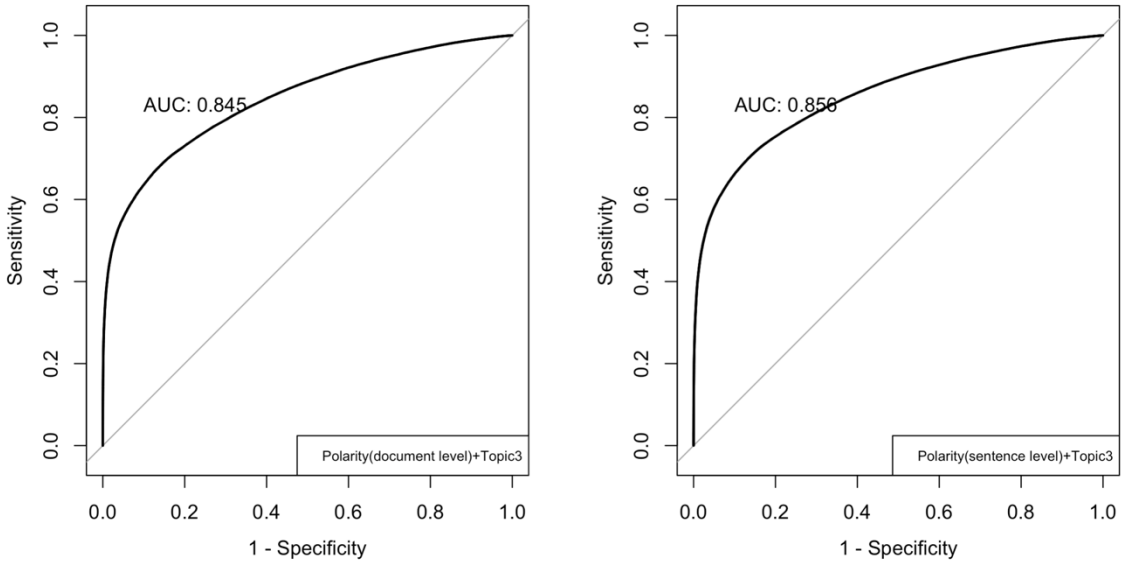

Fig. 7 ROC curves for Model D and Model E after tuning the hyperparameters of XGBoost

4.6 Rating score prediction

To examine the ability of polarity and topic membership to predict the exact rating score, we construct the baseline model, which includes each review's (sentence-level) polarity score as the only predictor. The sentence-level polarity will be adopted in the rating score prediction, considering that it is more accurate than document-level performance from the results in the previous section. Combining each topic membership separately with polarity score, 15 models are formed and trained using the same training dataset used in the binary classification task and performed using XGBoost. Considering that this task is an actual prediction task, MAE, and RMSE models are used for evaluation.

We calculate MAEs and RMSEs for 15 models as well as the baseline model. Fig. 8 displays the relative difference between MAE and RMSE for 15 models compared with the baseline model and sorts from the highest change to the lowest change. All topic membership could decrease the error compared with the baseline model, which indicates that the inclusion of even one topic as a covariate could increase the accuracy of the prediction task. There are two distinct variables (Topics #13 and #3) that can dramatically decrease the MAE and RMSE.

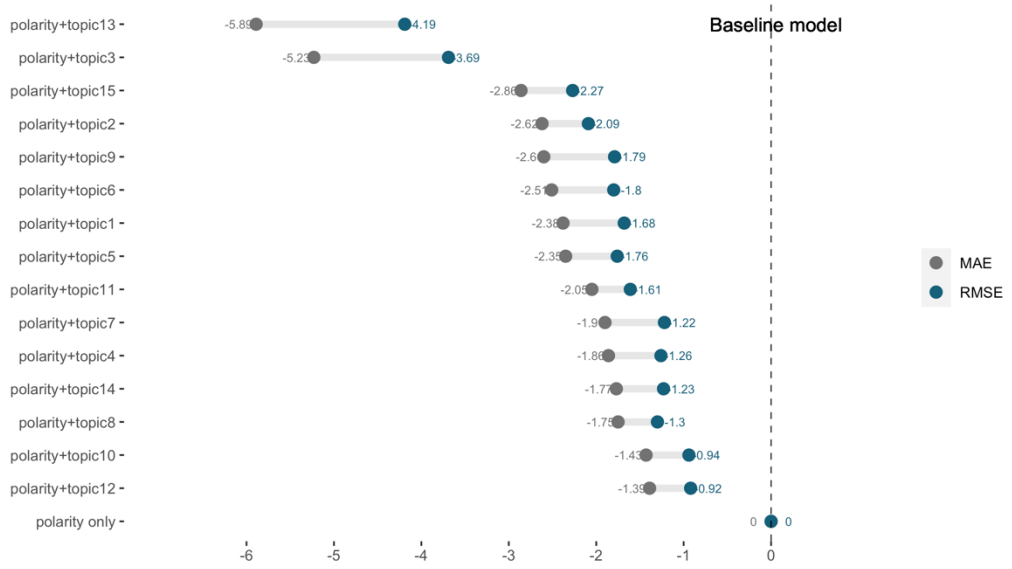

Fig. 8 Relative difference of MAE and RMSE for 16 models compared with the baseline model.

4.7 SHAP feature importance

Even Topic #13 and Topic #3 perform best in reducing the model error; other topics (i.e., Topics #15 and #2) also show improvement compared with the baseline model. To get a more specific and direct comprehension of how much each topic can contribute to the prediction of customer ratings, we construct a model including polarity score (sentence-level) and proportions of 15 topics as covariates to predict customer ratings. By including all topics, the accuracy is improved substantially (MAE=0.869, RMSE=1.137). The contribution of each

feature to the target value is represented by SHAP feature importance, which is calculated as the average of the absolute Shapley values for each feature across the dataset.

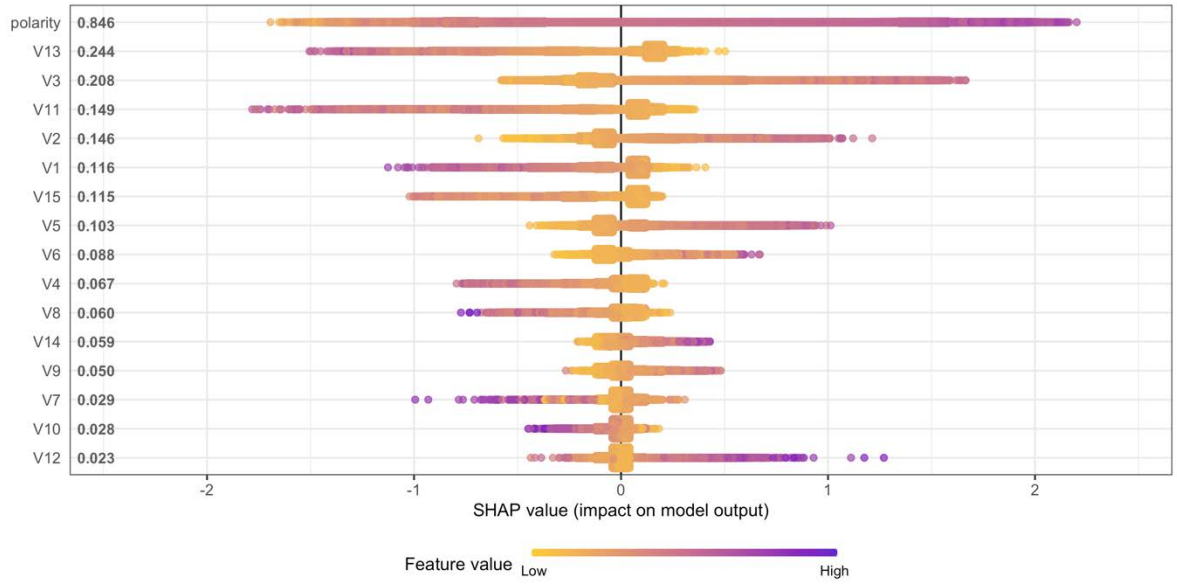

Fig. 9 SHAP summary plot

The summary plot (Fig. 9) displays global feature importance, as well as feature effects. All variables are sorted by decreasing feature importance along the y-axis, with their corresponding value next to them. The polarity score is the most dominant feature, and Topic #13 is the second most important feature, followed by Topic #3, while Topic #12 contributes the least to the predicted values. Each dot in this plot shows the Shapley value of an instance for each feature, whose horizontal location displays its Shapley value, and the specific feature determines the vertical location. The gradient colour demonstrates the original value for that variable from low to high. Polarity affects the target variable positively, as high polarity scores could increase the predicted customer ratings. Topic #13's membership is negatively associated with the target value as the predicted rating score will decrease while the proportion of Topic #13 within a review increases.

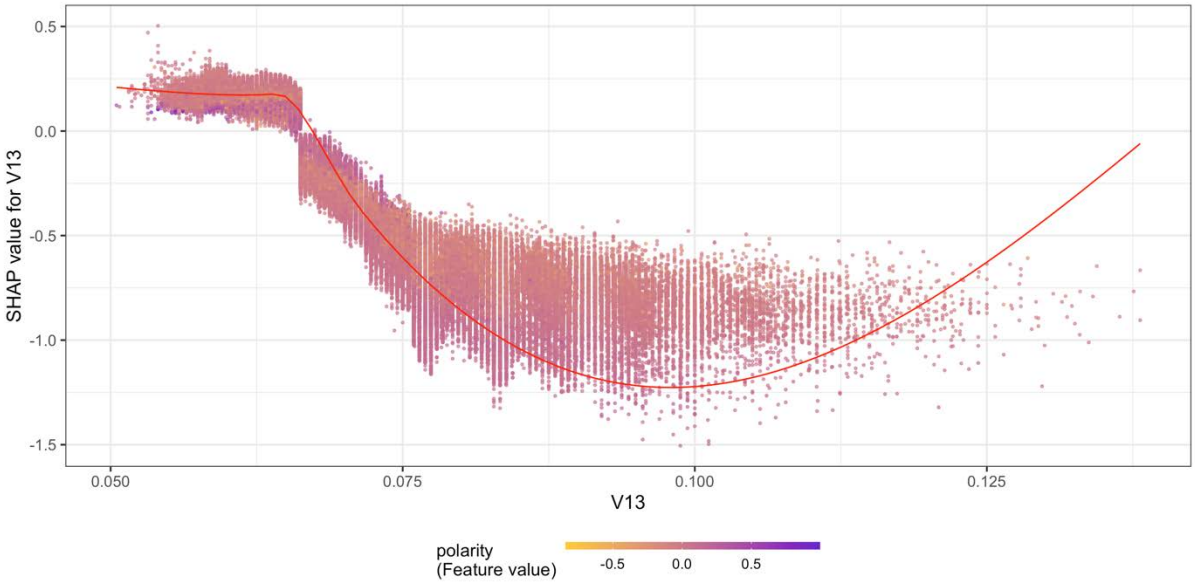

Fig. 10 SHAP dependence plot for Topic #13 membership and its interaction visualisation with polarity score

Fig. 10 displays the dependence plot for Topic #13 and its interaction with the polarity value. Each dot represents an instance with its proportion within a single review on the x-axis and its corresponding Shapley value on the y-axis. The gradient colour shows its polarity value. A small number of dots with proportions lower than 0.067, have positive SHAP values, indicating an increased prediction value. In contrast, a great many instances have a higher topic proportion between 0.067 and 0.096. Their corresponding SHAP values are lower than 0, meaning that they decrease the predicted value. More explicitly, for these dots whose x-axis is between 0.067 and 0.096, as the proportion of Topic #3 increases, their negative influence on the predicted rating score will be stronger.

5 Discussion and Implications

5.1 Discussion

Through the robust experiment (binary classification), the results show that both document (review)-level polarity score and sentence-level polarity score perform well in classifying a review as positive or negative with AUC scores—with both being higher than 0.8 even if they are included as the only covariate in two models, respectively. Compared with document-level polarity, sentence-level polarity performs better in the classification task, with a higher AUC score (0.838) than the other AUC score (0.827). Polarity within customer textual content could excellently predict customer satisfaction, while sentence-level polarity has a better ability for prediction, which identifies the impact of different granularity levels. It confirms the strong

ability of sentiment to explain customer ratings, consistent with Chatterjee et al. (2021) and Y. Zhao et al. (2019).

However, the topic memberships of 15 latent topics extracted from review text using a topic modelling approach (LDA) performed better in the classification task, with an AUC score of 0.860. Topic membership (all topics included) has higher accuracy than the classification task's polarity score (both document-level and sentence-level). It represents the multidimensionality in customer reviews in which they discuss their opinions towards the food and deliver service from various aspects. The multidimensionality is consistent with other studies which focused on customer reviews from other hospitality industries, such as airlines, restaurants, and hotels (Büschken & Allenby, 2016; Xu, 2020). In addition, the multidimensionality not only stands for various entities (e.g., food quality and delivery service) but also demonstrates customers' dialectic in the textual content. Customers might describe their experiences dialectically, through two-sided reviews (Wang et al., 2022). That may explain the stronger ability of topic membership to predict ratings, as both document-level and sentence-level sentiment could not capture the various dimensions (Birjali et al., 2021).

By examining each topic membership to the classification task separately, the one with the highest AUC score is selected and collaborated with the polarity score (two different levels), improving the AUC scores to 0.845 and 0.856, respectively, after hyperparameter tuning. The robust check could prove that the features generated from the textual content could be combined with the sentiment to achieve higher accuracy. It also reveals the heterogeneity of the latent dimensions (Büschken & Allenby, 2016; Hu et al., 2019), which do not equally contribute to customers' overall ratings.

Therefore, we included polarity (sentence-level) and each topic membership one by one as covariates and constructed 15 regression models. Compared with the baseline model (with only sentence-level polarity score), the results indicate that whichever topic membership could improve the model performance. Among them, two topics (Topic #13 and #3) could add the most accuracy to polarity in predicting rating scores since the MAE and RMSE of that model are decreased most compared to the baseline model. The information captured by the two topics could add more predictive accuracy to sentiment for rating prediction, which is also proved by the SHAP feature importance when we include all topics' membership and polarity (sentence-level) into the prediction.

5.2 Theoretical implications

Many studies have examined the power of sentiment and latent dimensions within review text to explain and predict overall customer satisfaction (Cheng et al., 2018; Xing et al., 2019; Y. Zhao et al., 2019). Compared with most previous research, our study reveals the comparison of the two approaches and how they can be combined for rating prediction.

First, our findings suggest that compared with sentiments, the topic memberships of latent dimensions generated from the review text have better performance in rating prediction. The top membership could be considered a helpful tool to predict customer overall satisfaction. Furthermore, the heterogeneity of the predictive power exists across different dimensions. Several dimensions have good performance in rating prediction, while several dimensions do not show good enough performance. The new features we extracted from the review text can be collaborated with customer sentiment to achieve better performance in rating prediction both holistically and individually. It extends the literature by combining the topic membership with customer sentiment instead of only adopting one type of feature from the review text.

Second, researchers have proposed several models to achieve higher accuracy in rating prediction (Cheng et al., 2018; Xing et al., 2019). Compared with these approaches, our approach has more flexibility and interpretability, especially when considering points of intervention in the customer-facing areas of the business. The use of the topic modelling approach allows us to extract latent dimensions from the textual data without pre-labelling the data, which saves human efforts in training and adjusting the model. The adoption of an unsupervised machine learning approach (LDA) saves the human effort in training the model. Besides, the adoption of Shapley value could clearly show how each dimension extracted from the review text contributes to predicting the overall ratings, which provides more interpretability of how each latent dimension and customer sentiment affect the overall customer satisfaction.

5.3 Practical implications

Customer ratings are direct measurements of overall customer satisfaction, while the textual content represents customer perception towards experience, showing customer satisfaction indirectly (predicting overall). Findings from this study could provide several managerial insights for restaurant owners and managers. First, our findings provide restaurants with the identification of latent dimensions from a large amount of customer reviews, demonstrating customers' various aspects of perception towards food and delivery service. Praise and customer complaints could help restaurant owners develop operational strategies.

Customer feedback is a vital information source for understanding customers' opinions, which has been proved to have a significant influence on customer behaviour and sales (Li et al., 2019; Z. Zhao et al., 2019). The 15 topics extracted from the review text show the multidimensionality of customers' evaluation. For instance, Topic #13 and Topic #8 illustrates customers' complaint about long delivery time and order issues, respectively, while Topic #2 represents customers' praise towards food quality.

Furthermore, heterogeneity exists in the contributions of each topic membership to customer overall satisfaction, which helps restaurants to prioritize the most influencing factors. By identifying the most important positive and negative dimensions that influence the rating scores, restaurant managers could explicit dissatisfaction and enhance their weaknesses accordingly or develop marketing strategies by highlighting their strengths. For instance, our findings show that except for polarity, Topic #13, mainly illustrating the long delivery time, contributes most to predicting the rating score. And the long delivery time is the most important factor negatively affecting the rating score as the lower membership leads to a higher rating score.

6 Conclusions, Limitations, and Future Research

Finding the underlying reasons for rating scores using contextual information is critical for businesses to develop strategies to discover why customers have different satisfaction levels. To discover the value of unstructured text within customer reviews to explain actual customer ratings, we evaluate and compare how two approaches (*sentiment analysis* and *topic models*) can be applied to understand customer satisfaction. This study demonstrates that incorporating document-level covariates, such as topic membership, can greatly contribute to the understanding of the sentiment of customer feedback, such as the ones found in customer reviews, and predict the review score in a much better way than the actual tone of the review text. While a large body of literature has demonstrated that review text is primarily consistent with the rating—and therefore, the sentiment of the review is reflected on the star rating of this particular review, from a business owner's point of view—latent dimensions discovered from these review texts are useful to be incorporated in the business practice. They can identify areas of improvement and competency the business can expand the most.

Nevertheless, there are several limitations to our research. For the robust check in our first experiment, we only classify reviews into positive and negative classes, not considering the neutral class, which has been commonly studied in online review literature. Furthermore,

even though we have examined two granularity levels of sentiment, we only employ a single dictionary. The choice of dictionaries may have a different influence on prediction accuracy. Therefore, these aspects could be improved in future work.

Future work should focus on several directions. First, reviews classified as neutral could be considered, together with the positive and negative ones, as three distinct classes for classifying customer rating scores to different satisfaction levels. Second, apart from lexicon-based methods, other unsupervised machine-learning techniques could be adopted to detect customers' polarity scores. Also, the subjectivity and emotions contained in the review text could be considered in future work. Third, customer reviews from other platforms could be included in the future to discover the level of heterogeneity across different platforms.

References

Al-Natour, S., & Turetken, O. (2020). A comparative assessment of sentiment analysis and star ratings for consumer reviews. *International Journal of Information Management, 54*, Article 102132. https://doi.org/10.1016/j.ijinfomgt.2020.102132

Alghamdi, R., & Alfalqi, K. (2015). A survey of topic modeling in text mining. *International Journal of Advanced Computer Science and Applications, 6*(1). https://doi.org/10.14569/IJACSA.2015.060121

Bagheri, A., Saraee, M., & De Jong, F. (2013). Care more about customers: Unsupervised domain-independent aspect detection for sentiment analysis of customer reviews. *Knowledge-Based Systems, 52*, 201-213.

Bastani, K., Namavari, H., & Shaffer, J. (2019). Latent Dirichlet allocation (LDA) for topic modeling of the CFPB consumer complaints. *Expert Systems with Applications, 127*, 256-271. https://doi.org/10.1016/j.eswa.2019.03.001

Birjali, M., Kasri, M., & Beni-Hssane, A. (2021). A comprehensive survey on sentiment analysis: Approaches, challenges and trends. *Knowledge-Based Systems, 226*, 107134.

Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM, 55*(4), 77-84. https://doi.org/10.1145/2133806.2133826

Blei, D. M., & Lafferty, J. D. (2007). A correlated topic model of science. *The Annals of Applied Statistics, 1*(1), 17-35.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research, 3*, 993-1022. https://doi.org/https://dl.acm.org/doi/10.5555/944919.944937

Brintrup, A. (2021). AI in the Supply Chain: a classification framework and critical analysis of current state. In *Oxford Handbook of Supply Chain Management*. OUP USA. https://doi.org/10.1093/oxfordhb/9780190066727.013.24

Büschken, J., & Allenby, G. M. (2016). Sentence-based text analysis for customer reviews. *Marketing Science, 35*(6), 953-975.

Cao, J., Xia, T., Li, J., Zhang, Y., & Tang, S. (2009). A density-based method for adaptive LDA model selection. *Neurocomputing, 72*(7-9), 1775-1781. https://doi.org/10.1016/j.neucom.2008.06.011

Chang, J., Gerrish, S., Wang, C., Boyd-Graber, J. L., & Blei, D. M. (2009). Reading tea leaves: How humans interpret topic models. In *Proceedings of the 22nd International Conference on Neural Information Processing Systems*, (pp. 288-296).

Chatterjee, S., Goyal, D., Prakash, A., & Sharma, J. (2021). Exploring healthcare/health-product ecommerce satisfaction: A text mining and machine learning application. *Journal of Business Research, 131*, 815-825.

Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, (pp. 785-794). https://doi.org/10.1145/2939672.2939785.

Cheng, Z., Ding, Y., Zhu, L., & Kankanhalli, M. (2018). Aspect-aware latent factor model: Rating prediction with ratings and reviews. In *Proceedings of the 2018 world wide web conference*, (pp. 639-648).

Deveaud, R., SanJuan, E., & Bellot, P. (2014). Accurate and effective latent concept modeling for ad hoc information retrieval. *Document Numérique, 17*(1), 61-84. https://doi.org/10.3166/DN.17.1.61-84

Dey, A., Jenamani, M., & Thakkar, J. J. (2018). Senti-N-Gram: An n-gram lexicon for sentiment analysis. *Expert Systems with Applications, 103*, 92-105.

Do, H. H., Prasad, P., Maag, A., & Alsadoon, A. (2019). Deep learning for aspect-based sentiment analysis: a comparative review. *Expert systems with applications, 118*, 272-299.

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters, 27*(8), 861-874. https://doi.org/10.1016/j.patrec.2005.10.010

Geetha, M., Singha, P., & Sinha, S. (2017). Relationship between customer sentiment and online customer ratings for hotels-An empirical analysis. *Tourism Management, 61*, 43-54.

Ghasemaghaei, M., Eslami, S. P., Deal, K., & Hassanein, K. (2018). Reviews' length and sentiment as correlates of online reviews' ratings. *Internet Research, 28*(3), 544-563. https://doi.org/10.1108/IntR-12-2016-0394

Ghiassi, M., & Lee, S. (2018). A domain transferable lexicon set for Twitter sentiment analysis using a supervised machine learning approach. *Expert Systems with Applications, 106*, 197-216.

Ghose, A., & Ipeirotis, P. G. (2011). Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer characteristics. *IEEE transactions on knowledge and data engineering, 23*(10), 1498-1512.

Griffiths, T. L., & Steyvers, M. (2002). A probabilistic approach to semantic representation. In *Proceedings of the Twenty-Fourth Annual Conference of Cognitive Science Society*, (pp. 381-386).

Guo, Y., Barnes, S. J., & Jia, Q. (2017). Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichlet allocation. *Tourism Management, 59*, 467-483.

Hofmann, T. (2001). Unsupervised learning by probabilistic latent semantic analysis. *Machine Learning, 42*(1), 177-196.

Hu, M., & Liu, B. (2004). Mining opinion features in customer reviews. In *Proceedings of the 19th national conference on Artifical intelligence*, (pp. 755-760).

Hu, N., Koh, N. S., & Reddy, S. K. (2014). Ratings lead you to the product, reviews help you clinch it? The mediating role of online review sentiments on product sales. *Decision Support Systems, 57*, 42-53.

Hu, N., Zhang, T., Gao, B., & Bose, I. (2019). What do hotel customers complain about? Text analysis using structural topic model. *Tourism Management, 72*, 417-426.

Kim, S., Park, H., & Lee, J. (2020). Word2vec-based latent semantic analysis (W2V-LSA) for topic modeling: A study on blockchain technology trend analysis. *Expert Systems with Applications, 152*, Article 113401.

Koltcov, S., Koltsova, O., & Nikolenko, S. (2014). Latent dirichlet allocation: stability and applications to studies of user-generated content. In *Proceedings of the 2014 ACM conference on Web science*, (pp. 161-165). https://doi.org/10.1145/2615569.2615680.

Kumar, A., Gopal, R. D., Shankar, R., & Tan, K. H. (2022). Fraudulent review detection model focusing on emotional expressions and explicit aspects: investigating the potential of feature engineering. *Decision Support Systems*, 113728.

Kwon, H.-J., Ban, H.-J., Jun, J.-K., & Kim, H.-S. (2021). Topic modeling and sentiment analysis of online review for airlines. *Information, 12*(2), 78.

Lai, X., Wang, F., & Wang, X. (2021). Asymmetric relationship between customer sentiment and online hotel ratings: the moderating effects of review characteristics. *International Journal of Contemporary Hospitality Management, 33*(6), 2137-2156. https://doi.org/10.1108/IJCHM-07-2020-0708

Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes, 25*(2-3), 259-284.

Li, X., Wu, C., & Mai, F. (2019). The effect of online reviews on product sales: A joint sentiment-topic analysis. *Information & Management, 56*(2), 172-184.

Liu, B. (2010). Sentiment analysis and subjectivity. In *Handbook of natural language processing* (Vol. 2, pp. 627-666).

Liu, B. (2012). Sentiment analysis and opinion mining. In *Synthesis Lectures on Human Language Technologies* (Vol. 5, pp. 1-167). https://doi.org/10.2200/S00416ED1V01Y201204HLT016)

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, (pp. 4768-4777).

Mai, L., & Le, B. (2021). Joint sentence and aspect-level sentiment analysis of product comments. *Annals of Operations research, 300*(2), 493-513. https://doi.org/10.1007/s10479-020-03534-7

Marshan, A., Kansouzidou, G., & Ioannou, A. (2020). Sentiment Analysis to Support Marketing Decision Making Process: A Hybrid Model. In *Proceedings of the Future Technologies Conference*, (pp. 614-626).

McAuley, J., & Leskovec, J. (2013). Hidden factors and hidden topics: understanding rating dimensions with review text. In *Proceedings of the 7th ACM conference on Recommender systems*, (pp. 165-172).

Minka, T. P., & Lafferty, J. (2002). Expectation-propagation for the generative aspect model. In *Proceedings of the Eighteenth conference on Uncertainty in artificial intelligence*, (pp. 352-359).

Molnar, C. (2020). *Interpretable machine learning: A guide for making Black Box Models interpretable.* Lulu.

Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up? Sentiment classification using machine learning techniques. In *Proceedings of the ACL-02 conference on Empirical methods in natural language processing*, (pp.). https://doi.org/10.3115/1118693.1118704.

Porteous, I., Newman, D., Ihler, A., Asuncion, A., Smyth, P., & Welling, M. (2008). Fast collapsed gibbs sampling for latent dirichlet allocation. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, (pp. 569-577).

Qiu, J., Liu, C., Li, Y., & Lin, Z. (2018). Leveraging sentiment analysis at the aspects level to predict ratings of reviews. *Information Sciences, 451*, 295-309. https://doi.org/10.1016/j.ins.2018.04.009

Qu, L., Ifrim, G., & Weikum, G. (2010). The bag-of-opinions method for review rating prediction from sparse text patterns. In *Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010)*, (pp. 913-921).

Quan, C., & Ren, F. (2014). Unsupervised product feature extraction for feature-oriented opinion determination. *Information Sciences, 272*, 16-28.

Rinker, T. (2020). *qdap: Bridging the Gap Between Qualitative Data and Quantitative Analysis*. R package version 2.4.3.

Roberts, M. E., Stewart, B. M., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S. K., Albertson, B., & Rand, D. G. (2014). Structural topic models for open-ended survey responses. *American Journal of Political Science, 58*(4), 1064-1082.

Seo, S., Huang, J., Yang, H., & Liu, Y. (2017). Interpretable convolutional neural networks with dual local and global attention for review rating prediction. In *Proceedings of the eleventh ACM conference on recommender systems*, (pp. 297-305).

Sharma, S. S., & Dutta, G. (2021). SentiDraw: Using star ratings of reviews to develop domain specific sentiment lexicon for polarity determination. *Information Processing & Management*, *58*(1), 102412.

Syed, S., & Spruit, M. (2017). Full-text or abstract? examining topic coherence scores using latent dirichlet allocation. In *2017 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, (pp. 165-174). https://doi.org/10.1109/DSAA.2017.61.

Tan, Y., Zhang, M., Liu, Y., & Ma, S. (2016). Rating-boosted latent topics: Understanding users and items with ratings and reviews. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*, (pp. 2640-2646).

Tirunillai, S., & Tellis, G. J. (2014). Mining marketing meaning from online chatter: Strategic brand analysis of big data using latent dirichlet allocation. *Journal of Marketing Research*, *51*(4), 463-479. https://doi.org/10.1509/jmr.12.0106

Verma, S., & Yadav, N. (2021). Past, present, and future of electronic word of mouth (EWOM). *Journal of Interactive Marketing*, *53*, 111-128.

Wang, Y., Zhong, K., & Liu, Q. (2022). Let criticism take precedence: Effect of side order on consumer attitudes toward a two-sided online review. *Journal of Business Research, 140*, 403-419.

Xiang, Z., Du, Q., Ma, Y., & Fan, W. (2017). A comparative analysis of major online review platforms: Implications for social media analytics in hospitality and tourism. *Tourism Management*, *58*, 51-65.

Xing, S., Wang, Q., Zhao, X., & Li, T. (2019). A hierarchical attention model for rating prediction by leveraging user and product reviews. *Neurocomputing*, *332*, 417-427.

Xu, X. (2020). Examining an asymmetric effect between online customer reviews emphasis and overall satisfaction determinants. *Journal of Business Research, 106*, 196-210.

Yadav, A., & Vishwakarma, D. K. (2020). Sentiment analysis using deep learning architectures: a review. *Artificial Intelligence Review*, *53*(6), 4335-4385.

Yeo, S. F., Tan, C. L., Kumar, A., Tan, K. H., & Wong, J. K. (2022). Investigating the impact of AI-powered technologies on Instagrammers' purchase decisions in digitalization era–A study of the fashion and apparel industry. *Technological Forecasting and Social Change*, *177*, 121551.

Zhang, H., Shen, F., Liu, W., He, X., Luan, H., & Chua, T.-S. (2016). Discrete collaborative filtering. In *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*, (pp. 325-334).

Zhang, W., & Wang, J. (2016). Integrating topic and latent factors for scalable personalized review-based rating prediction. *IEEE Transactions on Knowledge and Data Engineering*, *28*(11), 3013-3027.

Zhao, Y., Xu, X., & Wang, M. (2019). Predicting overall customer satisfaction: Big data evidence from hotel online textual reviews. *International Journal of Hospitality Management, 76*, 111-121.

Zhao, Z., Wang, J., Sun, H., Liu, Y., Fan, Z., & Xuan, F. (2019). What factors influence online product sales? Online reviews, review system curation, online promotional marketing and seller guarantees analysis. *IEEE Access, 8*, 3920-3931.

Appendix A Sample data

Table A1 Part of sample data (outcome variable in grey)

	Target Variable (Negative/Positive)	Review Text	Polarity (document level)	Polarity (sentence level)	Dominant Topic
1	**Negative**	ORDER AT YOUR OWN PERIL; food arrived cold, soggy, burnt and late. Ordered "gourmet" chips that turned out to be soggy cold oven chips. Ordered a breaded chicken fillet that was burnt black all over and soggy on the bottom, also cold, came with rice and salad that were spread all over the container, rice freezing cold salad brown. Everything else ordered was also cold and sub-par, definitly not restaurant food. Ordered the restaurant to complain and the owner had the cheek to deny all of the above, despite the fact that I had the food sitting right in front of me whilst I was talking to him. He refused a refund denying there was anything wrong with the food. I've just had to spend over an hour talking to a representative from just eat in order to get my £25 back. APPALLING FOOD & SERVICE!	-0.427	-0.194	Topic #1
2	**Negative**	Food arrived late and stone cold. Why? Because they used a cyclists to deliver the food. Which would you prefer, your food arriving quickly and hot by a moped delivery or slowly and cold by a cycle delivery? it's a no brainer, especially on one of the coldest evenings of the year! Food outlets, Stop using cyclists to deliver the food, it will inevitably arrive cold. It may be worth refusing the delivery if it does arrives by cycle. Food arriving late, cold and congealed is not my idea of a decent takeaway nor value for money! Stop this green nonsense and only use mopeds.	-0.248	-0.055	Topic #1
3	**Negative**	Food was stone cold, and chucked in the bag like you would never believe. They put the cold bottle of water in the same bag as the food, the food wasn’t in boxes the chips all fallen out in the bottom of the bag snd the chicken selects was all fallen out too. This made my bottle of water all greasy and the food stone cold from the cold bottle of water! What a night mare	-0.329	-0.141	Topic #1
4	**Positive**	Food was cold not happy about this as food was cold I got it from driver walked up my drive opend bag got my food out it was cold I	-0.270	-0.270	Topic #1

		didn’t order an ice cube sandwich wouldn’t use again due to food cold o did I say food was cold !!!			
5	**Positive**	The food was lovely, we ordered, 2 chicken kormas, rice, chips and peshwari naan. The curry was nice and plenty of chicken, the rice was standard too very tasty, the chips were plentiful and the peshwari was lovely!! Everything was very nice and hot on delivery, there were also complementary onion bhaji’s. The delivery was quick and the guy was really nice too. Definitely ordering from here again.	0.591	0.245	Topic #2
6	**Positive**	Lovely fresh, hot, tasty food. Lots of lovely fresh vegetables in dishes. Curry sauce nice and tasty. Very good value for such good quality food :) First time tried this; my new fave takeaway:)	0.758	0.402	Topic #2
7	**Positive**	Would order again. Food very tasty. Couple of minor issues - our son's rezala was listed as mild to medium, but had some very hot fresh chillies added which made it more like medium to hot and was too hot for him. The chana masala sauce was very nice, but chickpeas were quite hard. Seemed like they had been added late to the sauce, rather than cooked in it.	0.497	0.134	Topic #2
8	**Positive**	Ordered from here first time and will love to order again lamb chops so amazing and fresh and tasty, burger very tasty with fresh salad and wings and fries tasty and delicious also dessert chocolate concrete delicious everything was as hot and fresh as requested thank you	0.614	0.614	Topic #2
9	**Positive**	Hi there is Eric Clapton here again I can't believe this place are amazing food was nice hot and fresh everything was bang on 100% highly recommended this place no doubt to anyone thanks I'm in love with this place food was so delicious and plus I've phone in that carry out and the staff was excellent communication skill was excellent thanks thanks thanks	0.621	0.621	Topic #3
10	**Positive**	One of the best if not the best South Indian restaurants in London, I'm glad I'm within the area they deliver to. The tastes are authentic and their flavours are amazing , hats off to the chef for his amazing recipes and execution of the dishes. So far my favourites are the Nadan chicken fry,the pepper chicken, the fried beef and their non veg masala dosa. I'll be trying their chicken Ulathu and the Parrippu curry today. Hopefuy they will get a good review as well. I'm highly anticipating them to be as good as the previous dishes. These guys need to be supported for the fantastic food they produce and so it can reach née customer to try. Well done so far and I	0.428	0.151	Topic #3

		hope you guys keep consistent with the amazing taste. Wish you guys the best.			
11	**Positive**	Wow this place is amazing!! Order came early, staff are so polite and friendly, fantastic customer service and the slushies are out of this world!!! Taste like rainbow sherbet mmmm delicious! Everything was hot and fresh, perfect order 👌 definitely ordering again soon :) thanks guys	0.637	0.403	Topic #3
12	**Positive**	very very very very nice place to eat great service highly recommended very very very very tasty curry and half rice and chips and chicken wings one of my favourite takeaways keep up the good work great friendly and amazing service great food great place to eat excellent food excellent service excellent food and the best food in Manchester very very very very tasty food amazing service awesome food spot on	0.952	0.952	Topic #3
13	**Negative**	Delivered wrong order, telephoned them for nearly 30 mins before i got an answer, then said collect food from wherever it has been delivered, so expecting to get cold food as a minimum.. Just put phone down on me TERRIBLE SERVICE- TERRIBLE SERVICE -TERRIBLE SERVICE - TERRIBLE SERVICE -TERRIBLE SERVICE - TERRIBLE SERVICE - TERRIBLE SERVICE. No not sorted no food but less £16 can't even be bothered to call me back, perhaps customer service training would be a priority for them. NEVER ORDER FROM THEM AGAIN !!!!!!!	-0.415	-0.182	Topic #4
14	**Negative**	nobody at the restaurant will answer the phone. nobody at just eat will help. my order showed up without the tango i ordered. i don't want a refund i just want the tango i ordered. it seems no one wants to do their jobs properly. just send me the tango i ordered. bad service from just eat as when they do answer their phone they just put me through to an automated service on how they did. i used the key pad to give them a bad review only for the computer to tell me we didn't understand. poor service from everyone all round. just eat are no help at all when you try to call them. it sounds like i'm put through to an american who doesn't understand what the problem is. poor customer service	0.034	-0.029	Topic #4
15	**Positive**	Ordered 12 pack of mixed choice cookies, only received 10 on arrival. Contacted the restaurant to see where the other two cookies we paid for were, and the excuse was “We didn’t physically have them”. We were not contacted about this, and were told that we could come back and get the two we weren’t given but why should we? The lady on the phone was extremely rude, and told us to ring	-0.072	-0.022	Topic #4

		back tomorrow to speak to the manager as he/she was not in the building. Blamed it on the driver not waiting for the extra two cookies. When asked for the managers name, she hung up the phone without giving me an answer. I am extremely applaud at the service and we will be phoning back tomorrow to discuss this with the manager. We paid for 12 cookies, we should have received 12. Simple. No excuses and no apology given either. The lady needs to learn some manners and how to respect and talk to her customers with decency. She refused to give me her name as well, but will be sure to put in a complaint to the manager.			
16	**Negative**	I placed an Order with Food Fusion at 10:40pm and I received Luke Warm/Cold Food from Food Fusion at 12:20am. I am very Disappointed and Angry with Food Fusion. Their Lame Excuse is they do not have Drivers and their Drivers do not have Fuel in their Vehicle. The Whole of UK is in Crisis with Fuel Shortages and Panic Buying at Fuel Stations across the whole of United Kingdom. Other Restaurants and Takeaways are ensuring that their Drivers have Fuel in their Vehicle even if they need to Travel onto M11 and other Motorways to Purchase Fuel. Food Fusion should stop making Lame Excuses and find a way of providing Good Customer Service to all their Customers. The Customer Service I received from Food Fusion has been completely Rubbish. Food Fusion wants 6 Star Reviews when they are providing 1 Star Customer Service. This is unacceptable. If Food Fusion continue providing Rubbish Customer Service to their Customers they will lose Money as Customers will Purchase their Food at other Restaurants and Takeaways that are providing Good Customer Service. This is a lesson for Food Fusion to learn that the Customer is always right.	-0.364	-0.133	Topic #4
17	**Positive**	I mean, at this point I feel like I'm advertising Roosters! I've ordered so many times that I might as well just start working there lol! What a great food, when my 1,5 year old doesn't eat food I give, I quickly order from Roosters and problem solved! A healthy fast food restaurant! Always great food, always tasty, always hot and I don't know about others but I get earlier than presumed! Which is awesome! Thank you Roosters, I will be ordering more and more!	0.645	0.192	Topic #5
18	**Positive**	We order regularly from King Neptune and it's always a great experience. The food is excellent, always fresh and always hot, and the delivery is always on time. The guys at the shop are	0.614	0.216	Topic #5

		fantastic, always friendly and always deliver with a smile. Great food and great service, what more can you ask for??!!			
19	**Positive**	Always get our Indian meal from princess of India, everything is always spot on- its always hot always on time always delivered with a smile always reasonably priced (20%off) always, it's always delicious and it's always a weekly treat for our household with plenty to go around	0.284	0.287	Topic #5
20	**Positive**	Love the food. Eat here all the time and food is always great apart from the delivery being 5 minutes late this time they normally always deliver on time or early. Once again great food and great service, also recommend eating in the staff are great and always look after you	0.491	0.310	Topic #5
21	**Positive**	The food arrived hot and was well presented and packaged. The chicken was good quality and of a generous amount within a large pitta bread. The portion of fries were nicely cooked, although a very small portion (comparable to a 'regular / small fries' at a well known burger franchise). The salad was very small and basic. Being a very small amount of shredded iceberg lettuce, red onion and only two halves of a very thin slice of tomato. I would suggest to order a portion of salad with this meal, as I found that it had finished it with around 4 mouthfuls. The quality of the food is good overall, however I would suggest to this business that if they made some changes to their salad in portion size and quality, the meal would be tremendously improved. Well done on the quality of the meat and pitta though. Best wishes!	0.422	0.160	Topic #6
22	**Negative**	The food in this place is way too overpriced and it's not even good portions nor good quality. I bought sever things such as a classic hot dog for £5.95, it was small and looked like something you'd get for £2.50-£3. Naga chops for £12.95 and that wasn't even worth the price, came with a small amount of rice and 4 chops and NO REAL NAGA or anything and the portion size was small. Again after looking at the food, it definitely wasn't worth the price and again way too overpriced. Naga Wings, 4 pcs for £4.60 and it was dry and didn't come with anything and no real naga used. With that price from other restaurants, you'd get steak chips with it. The prices in MG GRILL is just way too overpriced and doesn't match the food quality and the portion sizes and it's a complete ripoff, I will never order anything from this place again and other customers should be careful.	-0.048	-0.006	Topic #6

23	**Negative**	Really disappointed in quality and standards of food. The clay chicken wings and buffalo wings look and taste the small. Really small. You get 10 for 6 plus quid. There's no taste just bland really tiny wings. The £5 plus beef burger tastes the same as the value £1.49 burger. Hard and dry tastes like a kebab. They just add extra sauce and salad yo make it big, but it tastes of nothing special. They gave a generous portion of sauce maybe to make up for late delivery which i appreciated and a free coke bottle considering i spent over £60 which i still appreciated. Im just really upset about the tiny bland wings which i paid so much for. I make better wings at home. Why the high price for tint wings. If it was outstandingly delicious i would've understood. The chips were nice. The paratha wrap is also very tiny considering you pay £5 plus for it. And its nothing outstanding, just average. Overall the only thing i enjoyed was the chips the sauces and the coke. Dont charge high prices for standard small portion food.	0.176	0.039	Topic #6
24	**Positive**	I have been ordering from here for about year now . the food taste great, however the prices don't much their portions. I am paying high end prices for food that don't even fill the Container. I am not asking for it to completely fill the container but at least be able to serve two people. I was extremely disappointed with my lastest order , the ribs was small and Bony, shredded salt and pepper chicken was extremely small. The egg fried rice was the only thing that came at a decent size, because everything else came small.I called to complain about the portion of the crispy shredded beef which cost £8.90 , to be honest I could have complained about most of the dishes portions but I chose to only complain about the crispy shredded beef because it costs me the most and I thought that was totally unacceptable for me to pay that price for that portion. after 5mins of explaining why that portion was unacceptable they Decided to send me other portion. I hope they Increase their portions for the future.	-0.176	-0.033	Topic #6
25	**Negative**	We ordered quite a lot of food. On this order there should be x2 nugget happy meals and x1 plain cheese burger happy meal but we recieved x1 plain cheese burger happy meal then inside the box was a book and an empty happy meal box that had not been assembled!! Not happy... luckily we ordered enough extras to make meals up for my children.	0.356	0.024	Topic #7
26	**Negative**	I ordered a Cheeseburger happy meal and a plain Burger happy meal and for some reason I've got a chicken nugget happy meal	0.044	0.065	Topic #7

		and A cheeseburger happy meal. Was not happy as my child does not like chicken nuggets and pointless contacting you guys because it only add credit to my account not giving me a new meal			
27	**Positive**	Food as expected, but not completely right. I ordered 1 nugget happy meal, 1 plain cheeseburger happy meal and 1 cheeseburger happy meal. I received 1 plain cheeseburger happy meal and 2 nugget happy meals. I had to give my eldest my food instead, because she doesn't like chicken nuggets.	0.144	-0.475	Topic #7
28	**Negative**	Ordered 2 Box meals for £18.50 including delivery. Was presented with 2 boxes with 4 pieces of chicken,a corn on the cob and pot of gravy and 2 cans...no fries.. the whole point of ordering a MEAL is to get a complete meal. Kfc have made the fries extra instead of including them as standard as they were previously. Every other fast food place when ordering a MEAL cones as a complete meal, main, fries and drink. Not here...you either have fries and no side or a side and no fries but then pay extra. Expensive treat for a disappointing small box of chicken. Kfc have taken away from the meals but not changed the prices to reflect that. Absolute con.	-0.046	-0.017	Topic #7
29	**Negative**	Absolutely awful service !! I placed an order via Whats App Heswall subway at 6pm , it should have arrived by 6:50pm . Received a text message to state order was on its way at 6:39pm. It's now 7:40pm and order still hasn't arrived. Rang Heswall where order was placed, to be told that the order has now gone to another store in Wirral for preparation, but they could not tell me which store. I was advised to search Google for telephone numbers of stores in wirral and try to locate the store myself to find where the order will be dispatched from !!!...whilst typing this review, the order finally arrived at 7:48pm, it arrived incorrect. so I queried the delivery driver as to where the rest of my missing order was ? I was advised all subway stores are shut and to ring up and complain tomorrow. Service is disgraceful !!! I now am out of pocket with an incomplete order and part of what did arrive was incorrect and the two sides that arrived out of the five sides that were ordered (3 missing) were stone cold . £27 wasted. Very Disappointing !!. Not deserving of any stars	-0.334	-0.152	Topic #8
30	**Negative**	My order was wrong. I was sent someone else's order! I called just eat who refused to send the correct order. Said they would send my email pictures to the investigation team and they would get back to me in 24 hours ! So i ended up with no order what so ever. I ordered 20 chicken nuggets , fries and a mayo chicken and got fries	-0.114	-0.079	Topic #8

		and a fish o fillet ! Mu daughter does not eat fish. Jennifer (clearly not her real name) from just eat point blank refused to send my correct order even though i emailed pictures of the wrong order. I am disgusted in this service and will be changing to another company in future . i spoke to the team at McDonald's who explained that was not right and my order should of been corrected right away. Still had no refund or my correct food ! No manners from just eat when we all know i paid for a service and didn't get what i paid for. Just eat your should be ashamed!!!!			
31	**Negative**	I ordered subway from Just Eat and it arrived completely wrong. Wrong size sub, wrong cheese, wrong toppings, wrong pops… just all wrong. I tried to call the restaurant but the number on just eat, Uber eats and google are all wrong. When I went to the website it wouldn't load. When I tried in app I couldn't get help. I finally found the right phone number in a phone book!!! I spoke to 'somu' he didn't want to give his name. He said he knew he messed up but it's coz he was busy. I asked him to send the correct items but he refused even though he knew the mistake he made. He told me to come to the restaurant and pick up the correct items. I told him I couldn't coz I'm home with two kids. He said he couldn't help me and said to come tomorrow! My kids are here without their tea! I'm very unhappy with this situation as I had to order from another place and I'm now out of pocket. His attitude was awful considering he knew the mistakes he made.	-0.367	-0.104	Topic #8
32	**Negative**	Wrong order to the wrong customer at the wrong address. Tried to contact restaurant to resolve the issue but the number given is 'invalid' which doesn't surprise me at all. I even asked the delivery driver if this was the correct order to which I was told it was what he was given. Eventually got the correct number (the last 2 digits are 79 not 97 as they listed) I got told to call up jusy eat and jump through a bunch of hoops to get the correct order.	-0.054	-0.063	Topic #8
33	**Negative**	I ordered a special chowmain, special fried rice, chicken noodle soup and szechuan chicken the special fried rice and chowmain where both okay but had little to no meat in them like a "special" variety of a meal should but they are a little bit cheaper than other takeaways so i suppose you get what you pay for, the chicken noodle soup was okay again not good but not bad they use the same noodles in it as the chowmain which is kindov weird ussually soups habe the thin noodles in them , now then the szechuan chicken was horrendous it was not like one i have ever had before im not sure	-0.124	-0.124	Topic #9

		they know what szechuan sauce is because this one was just like water with a few chillies in it like litterally the same consistency and taste of water, overall i dont think i could recommend it to anyone there was no real flavour to any of the food everything was very bland and underwhelming, there are alot of other options available to get Chinese from in Blackpool i sudjest you try them instead.			
34	**Positive**	We ordered butter chicken masala, butter paneer masala, saag paneer and pilau rice, expecting the style and flavours to be similar to many other places we have ordered these dishes from in the past. However, instead of a mild, creamy taste, each of the four dishes was excruciatingly hot and spicy with an overbearing flavour of star anise. All the dishes had this same burning flavour, and it wasn't an enjoyable dinner. If you like very spicy food it would be fine, and everything was a good temperature and well prepared and presented. It was just a bit of a shock to order what we have always thought were mild and creamy dishes and then not be able to eat them!	0.036	0.005	Topic #9
35	**Positive**	Beef in chilli sauce and Szechuan king prawns - both dishes looked exactly the same with a watery base, overflowing with vegetables and minimal portion of beef/prawn. Taste is almost the same as if pieces of boiled beef or prawn were added to a pre-prepared vegetable base. Crispy shredded beef tasted good. Egg fried rice was a little undercooked and minimal egg. Sweet and sour mixed vegetables - other than the sweet and sour taste and few pieces of additional vegetables again looked almost the same as the beef/prawn dish. Considering the overflowing vegetables in the beef and prawn, on hindsight, did not need the vegetable dish. Items should be renamed watery vegetables with few bits of beef/prawn. Sincerely sorry about this review, but unfortunately feel like money wasted.	-0.078	-0.011	Topic #9
36	**Positive**	We ordered the thai green chicken curry and the sweet and sour chicken. The thai green curry was watery and it didn't seem like coconut milk had been used (if it had it definitely was watered down). This meant there wasn't much flavour to the dish. My partner enjoyed the sweet and sour chicken but I found there was a strong after taste that didn't go away until I ate something else. I've had a fair few sweet and sour chicken dishes so I know that wasn't normal. I must say the appetisers were good and I couldn't fault them. Just a shame the main dishes were such a disappointment.	-0.029	0.010	Topic #9

37	**Negative**	Food was late barely warm called the restaurant got nothing but an attitude from staff on the phonentried to blame just eat for the food being late and barely warm definitely won't be ordering from this mcdonalds and will be putting a complaint to the mcdonalds complaint number	-0.199	-0.199	Topic #10
38	**Negative**	So I ordered at 17:30 (food was due at 18:15) called them at 18:30 was told they were waiting for the delivery driver.. called again at 19:00 still no driver called again at 19:30 and was asked if I could collect my food so I agreed, got a call back straight away was told the driver had arrived.. called again at 20:00 was told the driver hadn't arrived.. again, so I went and picked up my food, didn't even have the decency to apologise. Said I hope the food wasn't cold and was assured it was made fresh again, the food was freezing cold when I got home, tried to call again with no answer. Never again.	-0.047	-0.009	Topic #10
39	**Negative**	Originally ordered a gluten free sandwich and they sent a normal bread sandwich!!! Complained and ordered the same again and re-stated again, that I wanted gluten free. Guess what they did, they sent a non gluten free bread again but left a voicemail advising me its non gluten when the order was being delivered - how poor is that!! So twice got refunded and twice left hungry! Next time just cancel the order if you don't have gluten free bread Subway!!	0.055	0.036	Topic #10
40	**Positive**	I've tried all the big fast food restaurants during lockdown/Covid. This includes McDonald's and kfc. I've found burger king to be by far the best. Kfc give me the crappy food and I couldn't complain because I'm not there. And let's face it McDonald's food is bulk Cooke so even when I've used the drive through or just eat it's been luke warm. But today I gave Burger King a go as i love it anyway and I tell their didn't disappoint. Brilliant quality fresh and warm to eat. It's was definitely made to order I know it's a little more pricey but the food quality it's next to non. I eat a McDonald's and I'm Hungary again I two minutes. I'll never go any other than Burger King again. Well worth it!	0.201	0.098	Topic #10
41	**Negative**	Absolutely disgusting food. This is the second time I've ordered and after a long time since the last time it was appaulling disgusting food . The burger buns were rock hard the ribs smelt as if they wer weeks old the fried chicken is thee most horrible n dry chicken I've ever ate in my life . Do not order from this place they will serve you old dried bleachy smelling chicken shove it in a box with soggy chips fit n squeeze into one carrier bag n jus send it out to u	-0.246	-0.142	Topic #11

		it's absolutely the worst place ever . They do not fail to disappoint . Waste of money i couldn't eat any of it all of it went to the bin so disgusting that i wouldn't even feed it to my dog. It's vile. Do not waste your money on this absolute horrendous food.			
42	**Negative**	Really disappointed, the worst food I've ever ordered, what an absolute waste of money. Kebab overly salty, covered in old brown lettuce and chunks of cucumber - that's all the salad you get. Chicken dinner was two tiny dry drumsticks. Chips were so dry and hard it was like they had been reheated or something. Will not dare to order from there again ever. Don't have money to waste and wanted to treat my kids and it was absolutely disgusting. The dog ate the most of it.	-0.303	-0.067	Topic #11
43	**Positive**	We have ordered from here before and it has been fine. But this time, I have to say that it was absolutely disgusting. The pizza had a really hard base, it was awful, cheap, dough, it was inedible, my daughter couldnt even eat it. The chips were also old and hard and the fish was tasteless - it had no taste, with an old batter on it, and the chicken had hard crunchy coating on that also seemed old, it was the worst take away we have had, a waste of money. Sorry but thats the truth.	-0.421	-0.178	Topic #11
44	**Negative**	The worst Indian takeaway we have ever had! I have never ever written a negative review before but this was completely tasteless and we were disgusted- I couldn't even detect the taste of onion in the onion bhajis, or spices for that matter- just chewy and had a nasty taste of old oil. Saag aloo was TERRIBLE- again no taste at all, simply strange orange boiled wedges of potato in tasteless green spinach. I don't believe our garlic naan had any garlic on it though texture was good. Prawn passanda edible but not great, it is my favourite Indian dish and this was the worst one I've had. Bengal fish was simply dry white fish in plain tinned tomatoes. Tandoori chicken wings had an odd Chinese-style sauce. Mushroom rice was the only redeeming item we ordered as that was actually lovely and tasted good. Overall the worst curry we have ever eaten, most of it has gone in the bin- £30 wasted! It was also 30 mins later than the time we were given.	-0.068	-0.033	Topic #11
45	**Negative**	When I order I always say for delivery driver to press buzzer on main flat door. When I answer and say come in to my flat door they never do. They just wait at main door so I have to come out my flat and go to main door to get my order. I'm disabled and have trouble walking so that's why I always tell them to come to my flat door. I	-0.096	-0.038	Topic #12

		want just eat to tell there drivers to nit just stay at the main flats doors but to come to my flat door. It's not rocket science to know which is my flat door.			
46	**Positive**	The delivery driver tried to get me to ignore ‘contact free delivery’ to come outside to his car. I said no and to please deliver it contactlessly by delivering to my door. He said he was at number 37 (which is not my address and was no where to be seen outside). I explained that I live at the house with BLUE blinds. The driver then drove to a completely different house and left the food outside a random house, even though no other house on my street has my house number or colour blinds 🤦‍♀️ my husband then waited for him on the street and said the driver came up very close to his face. Unexpectedly poor delivery service this evening.	-0.091	-0.064	Topic #12
47	**Positive**	Delivery driver phones me asking where my address is located(like, don't you have a satnav or a map). As I was waiting at the door, a BMW pulls up on my road and bearing in mind this was at 2am, and during lockdown, so I assume it must be the delivery driver. He was in his car for a good 5minutes before phoning me asking where my house was. So I told him to get out of the car and cross the road. He spent another 5 minutes repositioning his car and opened his car door for a good minute before coming out and took his time to take out my order out of his car and to my doorstep. All this for £2.50 delivery charge and 50p service charge. Luckily by the time I tucked in to my food it was still warm. Saying that, the Food was great. And they gave us a 12" pizza rather than a 10". Thankyou. Also, Perhaps, the uncle delivery driver needs training in all aspects of this trade.	0.151	0.045	Topic #12
48	**Negative**	The driver did not follow the instructions I had given, we live on the ground floor of a small block of flats. I had described where our front door is and had asked that our food be delivered contact free and left outside our front door. The driver did not use the intercom to gain access to the block, or ring our doorbell. He rang to say that he was outside and had left our food outside our door as requested and checked the colour of our door. I opened the door but no food so I rang him back, he said "I don't know where you are I could only see 1 red door", so I told him I would come outside my door, then I realised he hadn't even attempted to enter the block and had left our food on the ground outside the block of flats. I told him I would let him in via the security intercom and to please deliver the	0.072	0.035	Topic #12

		food as I had asked. He eventually did this, but by the time we got the food it was cold. I have never had such bad service.			
49	**Negative**	Don't believe the order updates. Food arrived 1hr 10 minutes late. The order said it had been dispatched on time. 45 minutes later it still hadn't arrived so I called the restaurant. They said the order had left 6-7 minutes ago and would be with me in 20-25 minutes. 30 minutes later it still hadn't arrived and the restaurant stopped answering the phone yet the app said my order had arrived. Luckily the food arrived 10 minutes later. By that point the food was 1 hour 10 mins late. I appreciate it was a busy night but why say food had been sent out / delivered when it hadn't?!	0.049	0.000	Topic #13
50	**Negative**	My order was 10 minutes short of a 3 hours late....I had missing items off my order and it was all cold.....we had to heat it all up in the microwave. I had a message to say it was an hour late and its on its way. Then had another one to say sorry it's late an hour later which made it 2 hours late....finally we received it 3 hours late. I'm not normally on for moaning and giving the circumstances I understand the pressure but 3 hours isn't acceptable.	-0.052	-0.037	Topic #13
51	**Negative**	Food was ok, however it was an hour late. We received a text through Just eat 10 minutes after the expected delivery time of 1 hour advising the food was on its way, 30 minutes later still no food. We live 10 minutes drive from the Chippy! We phoned the business to ask where the food was to be told they had pressed the button advising the food was on its way by accident but the food was definitely on its way. 15 minutes later it still hasn't arrived 😠 When the food finally arrived it was an hour late. Doubt I will be ordering from here again, 2 hours to prep and deliver fish and chips possibly you shouldn't take more delivery orders than you can handle.	-0.082	-0.016	Topic #13
52	**Negative**	Two hours and half hours delivery it takes to deliver, with no communication or courtesy, called Just Eat to call them, mentioned 10mins time it will come and £3.50 refund for compensation and at this point it was one and half hours but 10 mins later again nothing, until one hour later calling Just Eat again even Just Eat was fed up with the take away so offered me £5 Just Eat Voucher on top for my next order, anyways got the food two and half hours later and the driver tells me he agrees waiting two and half hours is not right and the owner does not want to get another driver cos the second driver he had has gone Pakistan with no replacement.	0.045	0.025	Topic #13

53	**Positive**	Second time ordering first time was great & absolutely delicious. Second time we ordered 5 piece kids chicken nugget meals 3 lots and only come with 3 nuggets each and ordered salt n pepper chicken special no salad or sauce like last time and they called last time was looking forward but left a bit disappointed this time	0.054	0.136	Topic #14
54	**Negative**	I have ordered twice now to test it out. First time prompt delivery however not a busy evening. Second time was a Friday eve. I don't expect it fast however the app told me a change of times 4 times over whilst they assigned a driver. First time 1 out of 2 McFlurrys missing. Second time it asked for me to select 4 dip sauces for the family. None of them arrived. It's great to have it available however it needs fine tuning.	0.312	0.096	Topic #14
55	**Positive**	I have ordered 4 different times. First time food was 4 hours late and I was shocked, it arriving cold and I binned £25 worth of food! And second time it was on time but cold. Third time it arrived on time and not. Forth time it arrived on time and hot. All I can think is the first two times I order they were new on just eat and maybe had a lot of orders. Overall service is good. They never forget food which is good !	-0.055	0.042	Topic #14
56	**Positive**	second time ordering and it won't be the last! boyfriends first time trying the S&P box and it won't be his last time either!😂 both times have tasted unreal, same time next week😉	0.083	0.055	Topic #14
57	**Negative**	PUT THE DAMNED SAUCES IN THE BAG!! PUT THE DAMNED SAUCES IN THE BAG!! PUT THE DAMNED SAUCES IN THE BAG!! PUT THE DAMNED SAUCES IN THE BAG!! PUT THE DAMNED SAUCES IN THE BAG!! PUT THE DAMNED SAUCES IN THE BAG!! PUT THE DAMNED SAUCES IN THE BAG!! PUT THE DAMNED SAUCES IN THE BAG!! PUT THE DAMNED SAUCES IN THE BAG!! PUT THE DAMNED SAUCES IN THE BAG!! PUT THE DAMNED SAUCES IN THE BAG!! PUT THE DAMNED SAUCES IN THE BAG!! PUT THE DAMNED SAUCES IN THE BAG!! Seriously, it's really not that complicated, unless you're thick or lazy.	-0.594	-0.169	Topic #15
58	**Negative**	Multiple times now that drinks have not been secured properly in bags with lids not placed on properly causing them to spill in the bag and on other items and causing the bags to collapse underneath. Just today my hot drink had not been properly bagged meaning it	-0.236	-0.166	Topic #15

		spilt in to the bag causing the bag to collapse underneath itself and so the drink spilt all over the floor			
59	**Negative**	Only one drink not the full order, drink spilled all over the bag and dripped right the way throughout the house from the front door to the kitchen. Spilled drink in the same delivery thermal bag which has now ruined the full order i expect a full refund absolute joke	-0.071	-0.053	Topic #15
60	**Negative**	Drink had spilt ice the bag and was all over the food and in the bag. The chips was everywhere in the bag so was covered in drink and all soggy. There was ice in my drink when I asked for no ice but didn't matter as was swimming in the bottom of the bag. I only got to eat the burger just. Had no drink or chips.	0.000	0.000	Topic #15