Proposing a Systematic Approach for Integrating Traditional Research Methods into Machine Learning in Text Analytics in Tourism and Hospitality

Truc H. Le, Charles Arcodia, Margarida Abreu Novais, Anna Kralj
Department of Tourism, Sport and Hotel Management, Griffith University, Brisbane-Gold Coast, Australia

Truc H. Le *(corresponding author)*
Griffith University, 170 Kessels Rd, Nathan QLD 4111
Email: truc.le@griffith.edu.au

Charles Arcodia
Griffith University, 170 Kessels Rd, Nathan QLD 4111
Email: c.arcodia@griffith.edu.au

Margarida Abreu Novais
Griffith University, 170 Kessels Rd, Nathan QLD 4111
Email: m.abreunovais@griffith.edu.au

Anna Kralj
Griffith University, Parklands Dr, Southport QLD 4215
Email: a.kralj@griffith.edu.au

Proposing a Systematic Approach for Integrating Traditional Research Methods into Machine Learning in Text Analytics in Tourism and Hospitality

This paper argues that the analysis of vast amounts of user-generated content, which are currently dominated by text analytics and machine learning, need more methodical incorporation of reliable traditional methodologies to facilitate deeper understanding of concepts and theory building. Specifically, a systematic approach that integrates machine learning and traditional research methods is needed to overcome inherent drawbacks of both approaches. A step-by-step methodological framework for the analysis of online reviews is proposed and demonstrated. An application of the framework with an example drawn from the context of understanding authenticity in dining experiences illustrates its usefulness in the investigation of complex concepts. This paper represents one of the first attempts to systematise an integrated learning approach to understand complex concepts and build theories in tourism and hospitality, contributing to more rigorous procedures for processing and analysing large data sets of user-generated content.

Keywords

Machine Learning; Integrated Learning; Human Learning; Theory Building; Online Reviews; Authenticity

1. Introduction

Consumer behaviour has long been attracting substantial attention facilitated by various analytical methods with diverse sources of data. In particular, the growing power of electronic word-of-mouth have stimulated the prevalence of online customer reviews opening a myriad of possibilities for analysis in various areas of consumer behaviour (Mariani et al., 2018). Consumer behaviour, as Deighton (2007) asserts, despite being theoretically based, is also pragmatic in the sense that the research interest is often grounded between abstract theory and instrumental application. In tourism and hospitality in particular, not every consumer behaviour concept has received equal scholarly attention. On the one hand, various concepts have received limited attention triggering repetitive calls for scholarly action. On the other, further concepts have been subject to heightened debate, often creating more ambiguity than clarity, and ultimately rendering them too challenging to be operationalised robustly. As a result, developing advanced methodological approaches that can incorporate large amounts of textual data with conventional analyses can help tourism and hospitality researchers understand the use of language to inform consumer thoughts, and to develop more effective operationalisation of such concepts (Mariani et al., 2018).

Current analyses of vast amounts of user-generated content have relied on machine learning (Anandarajan et al., 2019). Machine learning is usually incorporated in text analytics to aid the detection of meaningful patterns and improve predictions based on that unstructured text data (Humphreys et al., 2018). Accelerated advancements in machine learning and text analytics have provided researchers with new approaches to collect, analyse and interpret this rich, unsolicited and arguably more authentic data source (Alaei et al., 2019; Li et al., 2018). Furthermore, tourism and hospitality industries are rooted and directed by consumer research, in which the methods are changing rapidly as capacities for collecting, storing, and analysing both textual and non-textual data have expanded (Ma et al., 2018). This inherent characteristic requires tourism and hospitality scholars to constantly seek powerful advanced research methods to examine abundant amounts of such heterogeneous and sophisticated data sources, and machine learning appears to be one of them (Alaei et al., 2019).

While having significant potential for drawing meaningful inferences from large amounts of data, machine learning is subject to several limitations, indicating there is still room for traditional research methods in understanding in depth a phenomenon (Le et al., 2020). Mariani et al. (2018) in their systematic review of big data applications in hospitality and tourism assert that it is still a challenge to complement small data from traditional methodologies with insights emerging from large amounts of user-generated content. Most machine learning studies requiring human-generated knowledge (apart from human input from supervised and semi-supervised learning) in tourism and hospitality thus far, have only incorporated traditional methodologies in an unsystematic and rather superficial manner that lacks methodical documentation and justification. Existing applications of data science in tourism and hospitality contexts need a progressive and methodical integration between the field expertise possessed by tourism and hospitality scholars, and the technical skills of computers and data scientists (Fuchs et al., 2013; George et al., 2016).

Traditional research methods have been widely used to collect and analyse a small amount of data that requires deep understanding of a phenomenon. Such data and the respective analytical methods suffer from substantial bias and errors that have been pertinent in social science research (Dolnicar, 2020; Mazanec, 2020). The large amount of readily available user-generated content, on the other hand, contains rich, unsolicited and arguably more authentic sources that better reflect consumer experiences and perceptions (Alaei et al., 2019), which leads to human inability to interpret such large amounts of data (Dey, 2016). Employing a systematic approach integrating traditional methods into machine learning therefore appears to compensate the limitations of both approaches in the analysis of vast amounts of user-generated content. Such analysis of vast amounts of data is far beyond the capabilities of some commonly used traditional content analysis software in social science research such as NVivo or Leximancer.

Moreover, some researchers are already claiming 'the end of theory' considering the 'data deluge makes the scientific method obsolete' (Andersson, 2008, p. 1) and data-led approaches render theoretical developments and explanations less important (Jackson, 2016). Theories nevertheless are still the ultimate prerequisite for consolidating and interpreting data. As a result, theories are and will always be required as the 'narrative way' behind knowledge generation (Mariani et al., 2018, p. 3542), which consistently encourages mixed-method approaches, triangulation and sense making through the use of theoretical frameworks in big data and machine learning studies (Mazanec, 2020). It is therefore imperative to combine the two learning techniques (i.e. traditional methodologies and machine learning) more transparently and systematically, especially for research that aims to gain a deeper understanding of concepts that potentially contributes to theoretical development.

This paper hereby proposes a systematic process that integrates machine learning and traditional research methods to analyse online reviews in tourism and hospitality contexts, with the aim of enhancing conceptual understanding and theory development of social phenomena. The paper first briefly outlines the strengths and limitations of machine learning, followed by an overview of machine learning studies in tourism and hospitality that incorporate traditional research methods and that aim to gain deeper conceptual understanding and theory development. The overview of existing literature highlights the significance of methodically integrating conventional methodologies into machine learning techniques to contribute to theoretical development of a phenomenon. To address this gap, a generic step-by-step methodological and analytical framework is presented to direct analysis of online reviews. The paper then illustrates an example based on understanding authenticity in dining experiences to exemplify the application of the integrated learning framework, following by comments directed towards what has been learnt and future directions.

2. Machine Learning in Text Analytics

Machine learning is generally defined as the construction of software that automatically detects meaningful patterns in given data to improve predictions and overcome human inability to interpret patterns or extract information from an abundance of data (Anandarajan et al., 2019). Despite the accelerated growth of machine learning, it is essential to understand its power in this era of digital information and in tourism and hospitality contexts particularly, as well as associated limitations which suggest some cautionary measures. Machine learning has strong potential to augment text mining capabilities, especially user-generated content analytics, which allows for the development of new knowledge to reinforce understanding of key concepts and to support decisions (Xiang et al., 2015). With the rise and the significant impact of user-generated content on understanding and driving consumer behaviour, machine learning capabilities aid the understanding of the concepts which emerge from this data source and predict future trends and patterns by continuing to acquire new data, which makes machine learning outperform traditional statistical methods in predictive power (Anandarajan et al., 2019; Jackson, 2016).

Machine learning, however, imposes several major limitations if used uncritically. Since diagnosing and correcting errors generated from the dataset is a challenging task, the step of cleansing and pre-processing data is essential to remove irrelevant data and increase the accuracy and comprehensibility of machine learning models (Kotsiantis et al., 2007). Data are a key ingredient that makes machine learning possible in that one can have machine learning without sophisticated algorithms but not without good data (Humphreys et al., 2018). This problem calls for a greater attention to understand the sample representativeness for the training data, as past performance is no guarantee of future results (Jackson, 2016). Machine learning also inherits low interpretability and explainability of different contexts. This is indeed a drawback of machine learning application particularly in the case of tourism and hospitality, since research in these contexts is heavily related to consumer behaviour and perception. This is illustrated through language as a reflection of culture: without having background knowledge and experience within the culture, it is impossible to understand the nuanced meanings conveyed by specific words (Chen et al., 2019; Xu et al., 2019). Since machine programs do not understand people's cultural values and norms, human-generated knowledge is essential to direct and reinforce machine learning outputs (Ansari et al., 2018; Brynjolfsson & McAfee, 2017).

In tourism and hospitality contexts, machine learning in text analytics has been mainly utilised to extract features and opinions embedded in online reviews (Alaei et al., 2019; Kirilenko et al., 2018), while very few studies have been conducted to enhance conceptual understanding or contribute to theory building. A few notable machine learning studies that seek to provider deeper insights into a particular concept include Duan et al. (2016), who investigated different dimensions in online reviews to measure hotel service quality and performance; Hu and Chen (2016), who explored and developed prediction models on hotel review helpfulness; Ma et al. (2018), who employed deep learning to examine the helpfulness of user-provided photos embedded in online hotel reviews; Xiang et al. (2018), who examined the reliability of online hotel reviews by developing a text classifier to predict travel purpose; and Xu and Li (2016), who identified and compared determinants of customer satisfaction and dissatisfaction among hotel types. However, while such concepts (i.e. service quality, review helpfulness, review reliability, determinants of customer satisfaction) are multi-dimensional, they have been well-established in terms of conceptualisation and operationalisation in the literature. Further, these concepts are not the sole domain of the tourism and hospitality fields and are rigorously investigated in other discipline contexts. Accordingly, there is a dearth of research that attempts to use machine learning to understand and conceptualise tourism and hospitality concepts, specifically those that are complex and multi-dimensional.

3. A Systematic Approach for Integrating Traditional Research Methods into Machine Learning

Integrated learning in this paper is concerned with knowledge and insights stemming from traditional data collection and analysis techniques, which are then used to complement and direct machine learning processes. These traditional methods can include customer surveys, interviews, focus groups, and manual and statistical content analysis (Mariani et al., 2018). The scope of this paper therefore does not include typical human inputs for machine learning approaches such as supervised and semi-supervised learning. Also, methodological approaches that only employ machine learning (and data/text mining) to analyse data collected from traditional methods (i.e. customer surveys) are also not considered as integrated learning (see Chiang, 2014; Fuchs et al., 2013; Law et al., 2011; and Zhang & Huang, 2015, in which survey data were analysed using data mining and machine learning).

To date, there have still been limited tourism and hospitality studies combining machine learning and traditional research methods to facilitate the analysis of online textual data. Table 1 offers a synopsis of how traditional methodologies have been integrated into machine learning in tourism and hospitality research to analyse such data. Particularly, traditional methods acted as an extraction tool to manually select terms related to attributes (Gan et al., 2017; Chen et al., 2019) and manually label reviews into positive and negative subsets (Fu et al., 2019; Xiang et al., 2017), with the common goal of establishing a list of keywords as a reliable foundation to be used in sentiment classifiers. Kirilenko et al. (2018) conducted a mixed-method study by using data from traditional surveys and online reviews to evaluate automated sentiment classifiers against human raters. Chen et al. (2019), Fu et al. (2019), and Zhang et al. (2016) aimed to improve the reliability of machine learning outputs by validating them through expert knowledge and applying systematic reliability check procedures. Another study focused on enhancing the result comprehensiveness by complementing the sentiment classification outcomes with the insights gleaned from unstructured interviews (Calheiros et al., 2017). It is important to note, however, that despite the attempt at deepening consumer perception understanding, Calheiros et al. (2017) barely discussed or integrated insights from the unstructured interviews into sentiment classification findings. The complementary role of human-generated knowledge in directing and facilitating machine learning is not well argued in this study and appears as a relatively minor element in the methodology.

[Table 1 near here]

Further, documentation that demonstrates systematic procedures of how machine learning has been augmented by traditional methods is lacking, and no evidence is offered that establishes the validity and reliability of such methods. Kirilenko et al. (2018) reported the randomly sampled survey dataset to compare with other automated sentiment classifiers; this however was scantly reported. It seems that although Gan et al. (2017) and Xiang et al. (2017) consider human-generated knowledge as a somewhat critical fundamental element to direct and facilitate subsequent automated text analyses, they did not report any methodical procedures to integrate traditional methodologies into machine learning. Zhang et al. (2016) followed systematic procedures to ensure reliability of small sample content analysis as a robust check for machine learning outputs. Fu et al. (2019) conducted manual annotation and follow-up discussion conducted by two annotators, while further measuring inter-annotator agreement to resolve differences during the process. Only Fu et al. (2019) and Zhang et al. (2016) reported diligently outcomes for the robust check.

A reverse combination has also occurred, where traditional methods have been augmented by big data analytics. Specifically, Shi et al. (2016) integrated semantic web and big data analytics into social scientific research (i.e. questionnaire design, structural equation modeling, and path analysis) in order to improve the effectiveness of modeling relationships of stakeholders of tourism intangible cultural heritage. In another instance, Xu et al. (2019) employed a multidimensional scaling analysis using machine learning algorithms to show the semantic structure of tourists' experience expressed of sharing accommodation, which was then examined in relation to guest satisfaction using conventional statistical analysis (i.e. factor analysis and regression analysis). As a result, there has been evidence that combining big(ger) data techniques with traditional research methods has been employed by both social science and computer data scientists with the aim of deepening knowledge of social phenomena (Mazanec, 2020). This integration however remains scattered and opaque because on the one hand, social science researchers, especially academics with a more traditional approach to data analysis, are still reluctant to familiarise themselves with the new analytical tools and the move towards data-led approaches, which may mean less use of theories, hypothesis formulation and testing based on significance levels (Jackson, 2016). On the other hand, computer and data scientists possess hard-skill expertise and may present methodologies and findings in a complex manner for non-data scientists to comprehend, thus creating a challenge for future studies that wish to employ similar approaches in other research scenarios.

The overview of existing research in machine learning applications in text analytics in tourism and hospitality has evidenced a lack of systematic documentation to combine traditional research methods with machine learning. George et al. (2016) emphasise the significance of creating multi- and inter-disciplinary research teams to integrate progressively and systematically social science and data science. It is hereby imperative to provide a methodical framework for analysing online reviews depicting the integration of machine learning and traditional methods in a more systematic way. The proposed framework will be applicable for studies employing integrated learning to enhance conceptual understanding and contribute to theoretical development of social science phenomena.

A Step-by-Step Methodological and Analytical Framework

A step-by-step methodological and analytical framework for the analysis of online reviews is presented in Figure 1. The depth of analysis is divided into two levels: the first level seeks to provide deeper insights into the concept and/or contextualise the concept; and the second level seeks to establish a conceptualisation for multi-dimensional concepts. Both levels of analysis employ integrated learning and each level consists of two milestones. For the ease of demonstration, the framework labels traditional research methods as "human learning".

[Figure 1 near here]

Specifically, the first level of analysis consists of Milestone #1 and #2. Milestone #1 specifically deals with identification of terms describing the concept of interest and its corresponding attributes. Integrated learning is utilised during Milestone #1 with *n* steps, *w*here *n* represents the complexity of the milestone and depends on how well-established the concept and its corresponding attributes have been in the existing literature. Since this type of framework is for the examination of review texts, the concept being analysed no doubt takes more time and effort to gain a comprehensive list of terms. Human-generated knowledge therefore plays a crucial role in directing and validating machine learning outputs since model training may generate several novel insights that need validation from field experts. The identification of corresponding attributes should be implemented after term identification for the key concept by following the same steps. Nevertheless, this level of identification is deemed less troublesome than the identification of the key concept, as corresponding attributes tend to be more commonly examined when it comes to online review analysis (e.g. restaurant/hotel attributes). Further, the corresponding attributes may not need integrated learning since their lexicons have been established. The outcomes of Milestone #1 include the final list of terms used to describe a concept and the corresponding attributes to each concept, which will be combined in Milestone #2.

Milestone #2, on the other hand, deals with the combination of concepts and its corresponding attributes (so-called pair combinations) to create meaningful insights and determine the relationship between the pairs or with other concepts. In this second milestone, these pairs can also be used to further test hypotheses with other concepts. The expected outputs of Milestone #2 are the pair combinations used to either contextualise the concept, or to determine the relationship between the pairs, or to test the relationship with other concepts. The majority of research using text analytics through online reviews ends at Milestone #2 since the research objective is typically to test relationships between concepts using corresponding attributes as bridging elements (see Shi et al., 2016; Xu et al., 2019).

The second level of analysis facilitates research that aims to establish conceptualisation for multi-dimensional concepts and consists of Milestone #3 and #4. These milestones are listed as *"if applicable"* since not every concept needs to be conceptualised, or reconceptualised, and this depends largely on the research objective. These additional milestones are suggested to propose, test, and confirm the training model, which is then utilised to conceptualise the concept. These milestones are similar to conventional statistical techniques that use factor analysis on survey samples to verify the structure of a concept. The key difference here, however, is the nature and the number of data used, thus making the techniques utilised to analyse such data ultimately different (Kotsiantis et al., 2007). Before this attempt, a proposed conceptual framework should be provided, which shapes and directs the development of training and test sets in Milestone #3. The training and test sets are typical examples of outputs generated from traditional research methods that are used to facilitate the machine learning process. Nevertheless, in social science research, it is vital to ensure the validity and reliability of such outputs by enhancing rigor of sampling strategies and data analysis as dealing with consumer research is not a straightforward matter. The training model in Milestone #4 is then tested and confirmed using the training/test set created in Milestone #3. The final outcome at the end of Milestone #4 is the trained model with sufficient evaluation metrics to confirm the conceptual framework. This conceptualisation can serve as a foundation for further operationalisation.

The exemplification of the proposed framework is now illustrated using a study of authenticity in dining experiences. This example highlights the framework's practicality in examining and conceptualising complex concepts, especially those concepts that are consumer-oriented and often expressed more strongly through written artefacts such as online reviews (Schuckert et al., 2015).

4. Framework Exemplification through Study of Authenticity in Dining Experiences

This section presents a demonstration of the proposed framework in the context of authenticity in dining experiences (see Figure 2). The dataset for this study consists of over one million online reviews scraped from Zomato Australia during July 2018. The study aims to enhance understanding of a complex concept (i.e. authenticity) and attempts to conceptualise its multi-dimensionality. The gaps identified in the extant literature of authenticity in dining experiences also reveal the possibilities of employing online user-generated content such as online restaurant reviews to enrich the understanding and conceptualisation of authenticity in dining experiences (Le et al., 2018, 2019). Through the analysis of authenticity perceptions embedded in online reviews, this study aims to make both some theoretical advancement to the complex connotations of authenticity, as well as a methodological contribution by using machine learning techniques to analyse a considerable amount of online textual data. This methodological contribution is significant because to date, there have been limited studies in tourism and hospitality that use machine learning to enhance understanding or conceptualise multi-faceted concepts (Le et al., 2020). The two levels of analysis divided into the four milestones (adapted from Figure 1) are as follows:

[Figure 2 near here]

4.1. First Level of Analysis

The first level of analysis seeks to gain a deeper understanding of authenticity and determine the restaurant attributes mentioned in authenticity judgments. This example utilises an existing list of authenticity terms (*Existing AU Terms*) created by Kovács et al. (2014) and O'Connor et al. (2017). The list of authenticity terms generated in these studies can be considered as a typical example of knowledge generated by traditional research methods (i.e. dictionary search, online surveys) (Kovács et al., 2014; O'Connor et al., 2017). However, since this list is not considered as exhaustive (as limited by traditional methods), it is vital to capture an extensive list of authenticity terms.

Milestone #1

Identify terms used to describe authenticity and restaurant attributes (RA) associated with authenticity terms (AU)

This milestone utilises integrated learning to identify terms used to describe authenticity and restaurant attributes associated with authenticity terms. The identification for authenticity terms and restaurant attributes consists of four identical steps. Table 2 then illustrates the milestone outputs generated from a small sample of review texts.

[Table 2 near here]

Authenticity Terms. Step 1 involves the manual identification of terms used to describe authenticity emerging from online reviews. Since there has been a readily available list of terms describing authenticity (created by Kovács et al., 2014 and O'Connor et al., 2017), Step 1 only identifies new authenticity terms (apart from *Existing AU Terms*) from the dataset. A purposive sample of online reviews was selected and browsed (i.e. only reviews that either contained the terms from *Existing AU Terms* or contained the indirect or underlying authenticity expressions were selected). This manual browsing was ended only when achieving saturation of the desired outcomes. For this example, the saturation was reached at 2000 reviews containing *Existing AU Terms*. The implication of this logic for purposive sampling of 2000 reviews is that there will be higher chances for identifying similar authenticity terms within the reviews that have already contained the existing ones. This step was conducted by an expert researcher in the field of authenticity and checked by three others. These *Potential New AU Terms* were reviewed in terms of context which was referred to in the review; each term was added to the keyword list for further evaluation and validation. For the concept that does not have an existing list of terms, Step 1 is most fundamental to create a "dictionary" of terms used as a foundation to direct the later machine learning. In this case, instead of employing the existing list, the researcher used their field expertise to manually select appropriate terms describing the key concept emerging from the purposive sample of reviews. In both cases, this step of "human learning" was fundamental in validating the machine learning findings because machine learning does not understand the contexts but only identifies the terms that are highly associated to each other irrespective of contexts.

In Step 2, the *Potential New AU Terms* were scanned across the entire dataset, and the top 10 most highly associated terms were generated for each existing authenticity term by training the semi-supervised learning Word2Vec model over the entire reviews. Word2Vec is an embedded word technique that can build a semantic understanding of the text in a corpus based on context (Li et al., 2018; Shuai et al., 2018), thus enabling the identification of similar terms for the given AU terms. The outcomes generated from training the Word2Vec model included *Misspelled and Derivatives of Existing AU Terms*, which takes too long to be detected via manual scanning; and *Words Used in Similar Contexts*.

In Step 3, the *Potential New AU Terms* identified in Step 1 were validated against the *Words Used in Similar Context* generated by the Word2Vec model in Step 2 (see Table 3). The reasoning behind this validity check is that if a new term is found to be related to *Existing AU Terms*, it should be statistically highly associated with *Existing AU Terms.* The cut-off point for the association scores to be considered as highly associated is .5 (Li et al., 2018). Since all terms in bold were highly associated statistically with *Existing AU Terms*, they were considered as *Verified New AU Terms*.

[Table 3 near here]

In Step 4, the *Verified New AU Terms* were combined with *Existing AU Terms* to create a list of more concise authenticity terms (*Final List of AU Terms*). Figure 3 illustrates a word cloud containing the *Final List of AU Terms* resulting from the integrated learning. The list was then used to scan across the dataset to detect corresponding restaurant attributes. However, if a concise list of terms describing the key concept is already available, this procedure is not needed (see Chen et al., 2019; Gan et al., 2017).

[Figure 3 near here]

Restaurant Attributes. The identification of terms used for restaurant attributes follows the exact same steps outlined for the authenticity terms. However, since restaurant attributes investigated in this study are mentioned in relation to authenticity terms in online reviews, only those reviews containing words on the *Final List of AU Terms* were chosen for the manual browsing in Step 1. This was done in order to minimise too many restaurant attributes being identified at a time causing confusion for the later pair combinations. Also, two rules were established to compromise between issues encountered when trying to identify which restaurant attribute is mentioned in relation to the authenticity term: (i) if two or more restaurant attributes appear in the same sentence with the same authenticity term, pick the restaurant attribute with the shortest distance to the authenticity term, (this was underpinned by Hsiao et al.'s (2017) and Chen et al.'s (2019) notion that suggested the closer the distance is between the two terms, the more highly they are associated); (ii) if the two restaurant attributes both have the same distance to the authenticity term, pick both, since it is impossible for the machine to determine the relevance of one term to the other if the distance between each pair is equal. This creation of rules demonstrates how human expertise can be added to guide the machine learning process to mitigate against the limitations inherent in its automated independent nature.

Steps 2 and 3 for restaurant attributes and authenticity terms were identical since restaurant attributes can be considered as another concept. In Step 4, the *Verified New RAs* (generated from Step 3) and *Predetermined Word List of RAs* (generated from Step 1) were then categorised into several groups. Although semi-supervised learning can also generate cluster and label outputs for the actual restaurant attributes, in order to categorise the actual restaurant attributes into meaningful groups, it requires substantial field expertise (Chen et al., 2019). The categorisation was conducted based on an extensive review of existing literature in restaurant attributes. The outcome generated from Step 4 was a table consisting of actual restaurant attributes classified in specific groups (*Groups of RAs with Word List*). There were eight groups of restaurant attributes identified in this study: Ambience & Atmosphere; Establishment Type; Ethnicity & Destination; Experience; Food & Drink; Other Customers; Restaurant Business; and Service. These groups of restaurant attributes were then utilised in the following milestones to create pairs containing authenticity term and group of restaurant attributes (*AU-RA pairs*).

Milestone #2

Create pairs containing authenticity term and group of restaurant attributes (AU-RA pairs)

In this milestone, actual restaurant attributes were searched in only those reviews containing authenticity terms. Before this was done, the reviews were split into sentences, and the authenticity terms found in that review were recalibrated to sentence level. Restaurant attributes were searched within sentences of the review that contains authenticity terms (rather than the entire review) to minimise the probability of too many restaurant attributes picked up at a time. A list of pairs containing authenticity term and group of restaurant attributes (*AU-RA pairs*) per sentence per review was stored. As shown in Table 2 as milestone outputs, for a given sentence, authenticity terms, restaurant attributes, groups of restaurant attributes, and *AU-RA pairs* are presented.

4.2. Second Level of Analysis

The second level of analysis seeks to conceptualise authenticity as a multi-dimensional concept. The multi-dimensionality encompasses *Authenticity of the Other/the Thing, Authenticity of the Organisation*, and *Authenticity of the Self* as proposed in Le et al. (2019). These three dimensions of authenticity have emerged from the review of existing literature in authenticity in dining experiences. Specifically, *Authenticity of the Other/the Thing* denotes object-related authenticity (Wang, 1999), in which attributes reflected from the object are associated with the otherness (e.g. ethnicity, culture, origin, history). *Authenticity of the Organisation* refers to the underlying producer/organisation who projects their own values and characteristics to be considered as authentic (Le et al., 2019). Lastly, *Authenticity of the Self* indicates the realisation of true self that may not be associated with the presence of any objects (Wang, 1999). Before Milestone #3 was conducted, a conceptual framework depicting these three dimensions of authenticity was required as an underpinning theoretical model (see similar works that employ predictive data-driven modeling in Jackson, 2016; Sun et al., 2016; Zhang & Huang, 2015).

Milestone #3

Create training and test sets by traditional methods: Classify AU-RA pairs into proposed dimensions of authenticity

Prior to developing training and test sets, it was proposed that different *AU-RA pairs* (from Milestone #2) can signify different dimensions of authenticity, which shape authenticity as a multi-dimensional concept. The training and test sets were developed and closely followed the conceptual framework for authenticity using a proportionate random sampling strategy. Table 2 demonstrates a sample of outputs for the training set, indicating that for every *AU-*

RA pair in a review sentence, the pair will signify at least one dimension of authenticity (i.e. 100% if signifying one dimension, 50% if signifying two dimensions). For example, “professional service”, “quirky service”, “inventive choices”, “sincere waiter”, “sincere business”, and “moral restaurant”, given the context of those particular review sentences, 100% signify *Authenticity of the Organisation.* On the other hand, “real vibe”, “authentic food”, “unique meat”, “modern varieties”, “quirky staff” 100% signify *Authenticity of the Other/the Thing*. Finally, “homey porridge” 50% signifies *Authenticity of the Other/the Thing* and 50% signifies *Authenticity of the Self;* “genuine bartender” 50% signifies *Authenticity of the Organisation* and 50% signifies *Authenticity of the Self.*

Milestone #4

Train the classification model to confirm the proposed dimensions of authenticity

Since the goal was to confirm the AU-RA pair classification, predictive modeling was employed which then contributed to the theoretical development of authenticity conceptualisation. Specifically, the training set from Milestone #3 was provided to train the classification model using one label and multi-label classification models (see Kotsiantis et al., 2007), and the training model was then tested using the test set. The test set was created separately to the training set to ensure the accuracy and reliability of the training model (Jackson, 2016; Zhang & Huang, 2015). The outcome of this milestone was to confirm the existence of proposed dimensions of authenticity, thus confirming the multi-dimensionality of authenticity in dining experiences.

5. Conclusions and Implications

While machine learning has been touted as a powerful research technique in text analytics across many disciplines (Anandarajan et al., 2019), there are limited studies using machine learning to gain better understanding of complex concepts in tourism and hospitality. This paper proposes a systematic approach that integrates traditional research methods into machine learning in the analysis of online reviews, with the goal to enhance conceptual understanding and theory building. By proposing a systematic approach, this paper also critiques the strengths and limitations of machine learning and suggests ways to overcome them using the example provided. This paper henceforth contributes to the literature in several ways.

First, the concise step-by-step methodological and analytical framework for the analysis of online reviews is to fulfil research that aims to enhance theoretical development of a concept based on written artefacts such as online reviews. The application of the proposed framework in the context of understanding authenticity in dining experiences highlights its utility in exploring and conceptualising complex and multi-dimensional concepts. As shown, traditional collection and analytical methods (as “human learning”) were added and systematically reported in several steps across the four milestones to direct and complement the utilisation of machine learning. The example demonstrates that traditional research methods were systematically added to fulfil four purposes: (1) to diagnose errors and detect ‘noise’ in data to improve accuracy before training the data; (2) to validate highly associated terms detected by machine learning that did not take into consideration the context differences; (3) to aid the decision-making of important outputs that requires substantial field expertise which has not been mapped in machine learning; and (4) to direct machine learning process (i.e. by creating a training/test set) and overcome limitations generated by the automated independent nature of machine learning. On the other hand, the substantial drawbacks of traditional methodologies such as the inability to detect misspelled and derivatives of terms, the inability to determine highly associated terms which emerged from the large dataset, and the inability to validate hypothetical assumptions with a vast amount of online data have been supplemented by the capabilities of machine learning.

Second, this paper calls for greater attention to well-documented and more systematic integrated learning approaches of text analytics. By doing so, this paper has reaffirmed that there is always room for improvement in human knowledge about complex multi-dimensional phenomena, especially when there is a need for further validation and observation from other sources of data to understand the phenomenon fully. This paper is a preliminary effort to make machine learning applications more approachable in tourism and hospitality text analytics, and the proposed framework offers a useful starting point from which to develop more effective integration of traditional research methods and big data analytics. Future research may also consider applying this proposed framework to improve conceptualisation of complex concepts in tourism and hospitality, and employing integrated learning to further aid operationalisation of such concepts.

More importantly, while utilising a systematic approach for integrated learning results in a better understanding of complex concepts, theoretical advancement also offers practical insights for industry (Pan & Fesenmaier, 2006; Ullah et al., 2016; Xiang et al., 2007). For example, understanding the language representation of restaurants in the online domain helps make inferences about consumer perceptions and preferences regarding restaurant attributes, which in turn assists the development of online recommender systems (Xiang et al., 2007). Firms can develop long-lasting relationships with consumers once they understand the meanings that shape consumer perception towards a business brand (Lee & Bradlow, 2011). Knowing consumers' perceptions of a product's features could contribute to better describing the feature during product development and the design of marketing and communication strategies (Varela et al., 2013). In relation to the example in this paper, proposing and testing the conceptualisation of authenticity helps convey dimensions of authenticity that attract more customer attention that is of great importance for restauranteurs to focus on. By identifying underlying authenticity dimensions expressed through online reviews, the findings derived from the example will offer strategies to trigger and enhance perceptions of authenticity using authenticators (which are attached with important restaurant attributes found in the case) and well-attended dimensions of authenticity.

6. References

Alaei, A. R., Becken, S., & Stantic, B. (2019). Sentiment Analysis in Tourism: Capitalizing on Big Data. *Journal of Travel Research, 58*(2), 175-191.

Anandarajan, M., Hill, C., & Nolan, T. (2019). *Practical Text Analytics: Maximizing the Value of Text Data* (Vol. 2). Cham: Springer International Publishing AG.

Andersson, C. (2008). The end of theory: the data deluge makes the scientific method obsolete. *Wired Magazine, 16*(7), 1-3.

Ansari, F., Erol, S., & Sihn, W. (2018). Rethinking Human-Machine Learning in Industry 4.0: How Does the Paradigm Shift Treat the Role of Human Learning?. *Procedia Manufacturing, 23*, 117-122.

Brynjolfsson, E., & McAfee, A. (2017). The business of artificial intelligence: What it can and cannot do for your organization. *Harvard Business Review*, 1-2.

Calheiros, A.C., Moro, S., & Rita, P. (2017). Sentiment Classification of Consumer-Generated Online Reviews Using Topic Modeling. *Journal of Hospitality Marketing & Management, 26*(7), 675-693.

Chen, M.-C., Hsiao, Y.-H., Chang, K.-C., & Lin, M.-K. (2019). Applying big data analytics to support Kansei engineering for hotel service development. *Data Technologies and Applications, 53*(1), 33-57.

Chiang, W.-Y. (2014). Applying data mining with a new model on customer relationship management systems: a case of airline industry in Taiwan. *Transportation Letters, 6*(2), 89-97.

Deighton, J. (2007). From the Editor: the Territory of Consumer Research: Walking the Fences. *Journal of Consumer Research*, *34*(3), 279–282.

Dolnicar, S. (2020). Why quantitative papers based on primary data get desk-rejected by annals of tourism research. *Annals of Tourism Research, 83*, 102981.

Duan, W., Yu, Y., Cao, Q., & Levy, S. (2016). Exploring the Impact of Social Media on Hotel Service Performance: A Sentimental Analysis Approach. *Cornell Hospitality Quarterly, 57*(3), 282-296.

Fu, Y., Hao, J.-X., Li, X., & Hsu, C. H. C. (2019). Predictive Accuracy of Sentiment Analytics for Tourism: A Metalearning Perspective on Chinese Travel News. *Journal of Travel Research, 58*(4), 666-679.

Fuchs, M., Abadzhiev, A., Svensson, B., Höpken, W., & Lexhagen, M. (2013). A knowledge destination framework for tourism sustainability. *Tourism, 6*(2), 121-148.

Gan, Q., Ferns, B.H., Yu, Y., & Jin, L. (2017). A Text Mining and Multidimensional Sentiment Analysis of Online Restaurant Reviews. *Journal of Quality Assurance in Hospitality & Tourism, 18*(4), 465-492.

George, G., Osinga, E., Lavie, D., & Scott, B. (2016). Big data and data science methods for management research. *Academy of Management Journal, 59*(5), 1493-1507.

Hsiao, Y.-H., Chen, M.-C., & Liao, W.-C. (2017). Logistics service design for cross-border E-commerce using Kansei engineering with text-mining-based online content analysis. *Telematics and Informatics*, *34*(4), 284-302.

Hu, Y.-H., & Chen, K. (2016). Predicting hotel review helpfulness: The impact of review visibility, and interaction between hotel stars and review ratings. *International Journal of Information Management*, *36*(6), 929-944.

Humphreys, A., Wang, R. J.-H., Fischer, E., & Price, L. (2018). Automated Text Analysis for Consumer Research. *Journal of Consumer Research, 44*(6), 1274-1306.

Jackson, S. (2016). Prediction, explanation and big(ger) data: a middle way to measuring and modelling the perceived success of a volunteer tourism sustainability campaign based on 'nudging'. *Current Issues in Tourism, 19*(7), 643-658.

Kirilenko, A.P., Stepchenkova, S.O., Kim, H., & Li, X. (2018). Automated Sentiment Analysis in Tourism: Comparison of Approaches. *Journal of Travel Research*, *57*(8), 1012-1025.

Kotsiantis, S.B., Zaharakis, I.D., & Pintelas, P.E. (2007). Machine learning: a review of classification and combining techniques. *Artificial Intelligence Review, 26*(3), 159-190.

Kovács, B., Carroll, G.R., & Lehman, D.W. (2014). Authenticity and Consumer Value Ratings: Empirical Tests from the Restaurant Domain. *Organisation Science*, *25*(2), 458-478.

Law, R., Rong, J., Vu, H. Q., Li, G., & Lee, H. A. (2011). Identifying changes and trends in Hong Kong outbound tourism. *Tourism Management, 32*(5), 1106-1114.

Le, T. H., Arcodia, C., Kralj, A., & Abreu Novais, M. (2018). Proposing an Exploration of Authenticity and its Key Determinants in Dining Experiences. *Paper presented at CAUTHE 2018: Get Smart: Paradoxes and Possibilities in Tourism, Hospitality and Events Education and Research held February 2018 at Newcastle Business School*. The University of Newcastle.

Le, T.H., Arcodia, C., Abreu Novais, M., & Kralj, A. (2019). What We Know and Do Not Know about Authenticity in Dining Experiences: A Systematic Literature Review. *Tourism Management, 74*, 258-275.

Le, T. H., Arcodia, C., Abreu Novais, M., & Kralj, A. (2020). Using combined-learning to conceptualise constructs in tourism and hospitality. *Paper presented at the Paper presented at CAUTHE 2020: 20: 20 Vision: New Perspectives on the Diversity of Hospitality, Tourism and Events, Auckland, New Zealand.*

Lee, T.Y., & Bradlow, E.T. (2011). Automated Marketing Research Using Online Customer Reviews. *Journal of Marketing Research*, *48*(5), 881–894.

Li, W., Zhu, L., Guo, K., Shi, Y., & Zheng, Y. (2018). Build a Tourism-Specific Sentiment Lexicon Via Word2vec. *Annals of Data Science*, *5*(1), 1-7.

Ma, Y., Xiang, Z., Du, Q., & Fan, W. (2018). Effects of user-provided photos on hotel review helpfulness: An analytical approach with deep leaning. *International Journal of Hospitality Management, 71*, 120-131.

Mariani, M., Baggio, R., Fuchs, M., & Höepken, W. (2018). Business intelligence and big data in hospitality and tourism: a systematic literature review. *International Journal of Contemporary Hospitality Management, 30*(12), 3514-3554.

Mazanec, J. A. (2020). Hidden theorizing in big data analytics: With a reference to tourism design research. *Annals of Tourism Research, 83*, 102931.

O'Connor, K., Carroll, G.R., & Kovacs, B. (2017). Disambiguating authenticity: Interpretations of value and appeal. *PLoS One*, *12*(6), e0179187.

Pan, B., & Fesenmaier, D.R. (2006). Online information search: vacation planning process. *Annals of Tourism Research*, *33*(3), 809–832.

Schuckert, M., Liu, X., & Law, R. (2015). Hospitality and Tourism Online Reviews: Recent Trends and Future Directions. *Journal of Travel & Tourism Marketing*, *32*(5), 608-621.

Shuai, Q., Huang, Y., Jin, L., & Pang, L. (2018). Sentiment Analysis on Chinese Hotel Reviews with Doc2Vec and Classifiers. *Paper presented at the 2018 IEEE 3rd Advanced Information Technology, Electronic and Automation Control Conference, Chongqing, China*.

Sun, P., Cárdenas, D. A., & Harrill, R. (2016). Chinese Customers' Evaluation of Travel Website Quality: A Decision-Tree Analysis. *Journal of Hospitality Marketing & Management, 25*(4), 476-497.

Ullah, R., Amblee, N., Kim, W., & Lee, H. (2016). From valence to emotions: Exploring the distribution of emotions in online product reviews. *Decision Support Systems*, *81*, 41–53.

Varela, P., Ares, G., & Fiszman, S. (2013). Texture and Semantics: The Conceptual Structure in Consumers' Minds. *Journal of Sensory Studies*, *28*(3), 194-204.

Wang, N. (1999). Rethinking authenticity in tourism experience. *Annals of Tourism Research, 26*(2), 349-370.

Xiang, Z., Kim, S.-E., Hu, C., & Fesenmaier, D.R. (2007). Language representation of restaurants: Implications for developing online recommender systems. *International Journal of Hospitality Management*, *26*(4), 1005-1018.

Xiang, Z., Du, Q., Ma, Y. and Fan, W. (2018). Assessing reliability of social media data: lessons from mining TripAdvisor hotel reviews. *Information Technology & Tourism*, *18*(1), 43-59.

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.

Xiang, Z., Schwartz, Z., Gerdes, J. H., & Uysal, M. (2015). What can big data and text analytics tell us about hotel guest experience and satisfaction? *International Journal of Hospitality Management, 44*, 120-130.

Xu, F., La, L., Zhen, F., Lobsang, T., & Huang, C. (2019). A data-driven approach to guest experiences and satisfaction in sharing. *Journal of Travel & Tourism Marketing, 36*(4), 484-496.

Xu, X., & Li, Y. (2016). The antecedents of customer satisfaction and dissatisfaction toward various types of hotels: A text mining approach. *International Journal of Hospitality Management, 55*, 57-69.

Zhang, C., & Huang, Z. (2015). Mining Tourist Motive for Marketing Development via Twice-Learning. *Applied Artificial Intelligence, 29*(2), 119-133.

Zhang, X., Wu, H., Xu, A., Yu, Y., Li, H., & Lin, Z. (2016). Sentimental interplay between structured and unstructured user-generated contents. *Online Information Review, 40*(1), 119-145.

Table 1. Empirical Studies that Integrate Traditional Research Methods into Machine Learning in Text Analytics.

Author	Traditional Research Method	Purpose of the Method	Role of the Method
Calheiros et al. (2017)	Unstructured interview to obtain perceptions on the customer feedback and the hotel's main strategy	To provide in-depth knowledge to strengthen discussion of the results	As a qualitative element to complement machine learning outputs in a mixed-methods design
Chen et al. (2019)	Manually identify Kansei words and service characteristics; each service concept is assigned to the ten hotel service constructs based on expert knowledge	To ensure the correctness of Kansei words and service characteristics constructed by text mining	As a basis for directing text extracting and mining
	Hotel managers and customers are invited to determine if the generated words and service characteristics can express consumers' feelings or hotel service properties	To validate Kansei words and service characteristics generated by text mining	As a robust check for reliability of machine learning outputs
Fu et al. (2019)	Manually annotate sampled news articles to assign news sentiment polarities based on codebook guidelines; resolve annotation differences through follow-up discussions between the two annotators; measure inter-annotator agreement to determine the internal consistency; employ stratified sampling and randomly select 1,000 articles (i.e., 500 positive and 500 negative) from the annotated samples to build the benchmark corpus	To identify implicit sentiments that are often written from a third-person point of view, which can lead to confusion between negative and positive sentiments	As a training/test set for machine learning
Gan et al. (2017)	Manually identify words and terms related to restaurant	To create a reliable set of keywords for restaurant	As a training/test set for machine learning

	attributes; words with high frequency of use were then added into the expert-generated word lists	attributes which were then used as the sentiment classifiers	
Kirilenko et al. (2018)	Manually classify the datasets	To compare the performance of human raters with other automated sentiment classifiers	As a contestant of machine rating (to determine the effectiveness of human rating and machine rating)
Xiang et al. (2017)	Manually label reviews into a positive and a negative set	To generate a list of words which were then used as the sentiment classifiers	As a training/test set for machine learning
Zhang et al. (2016)	Manually code textual comments (manual content analysis)	To lessen the limits of machine learning during sentiment analysis	As a robust check for reliability of machine learning outputs

Table 2. A Sample of Outputs for Milestones #1, #2, and #3.

	MILESTONE #1 Goal #1: Identify Authenticity terms Goal #2: Identify and Categorise Restaurant Attributes			**MILESTONE #2 Goal: Identify Authenticity-Restaurant Attributes Pairs**	**MILESTONE #3 Goal: Classify Pairs into Authenticity Dimensions**		
Review Sentence	**AU Term**	**RA**	**Group of RA**	**AU Term - Group of RA**	**Authenticity of the Other/ the Thing**	**Authenticity of the Organisation**	**Authenticity of the Self**
…	professional	service	Service	professional-Service	0	100	0
…	real	vibe	Ambience & Atmosphere	real- Ambience & Atmosphere	100	0	0
…	real	chicken	Food & Drink	real- Food & Drink	100	0	0
…	inventive	choices	Food & Drink	inventive- Food & Drink	0	100	0
…	unique	meat	Food & Drink	unique- Food & Drink	100	0	0
…	unique	places	Establishment Type	unique- Establishment Type	100	0	0
…	unique	reputation	Image	unique-Image	100	0	0
…	homey	porridge	Food & Drink	homey- Food & Drink	50	0	50
…	authentic	food	Food & Drink	authentic- Food & Drink	100	0	0
…	quirky	staff	Service	quirky- Service	100	0	0
…	sincere	waiter	Service	sincere- Service	0	100	0
…	sincere	business	Restaurant Business	sincere- Restaurant Business	0	100	0
…	genuine	bartender	Service	genuine- Service	0	50	50
…	authentic	food	Food & Drink	authentic- Food & Drink	100	0	0
…	moral	restaurant	Establishment Type	moral- Establishment Type	0	100	0
…	modern	varieties	Food & Drink	modern- Food & Drink	100	0	0
…	modern	varieties	Food & Drink	modern- Food & Drink	100	0	0
…	quirky	service	Service	quirky-Service	0	100	0

Signifies *Authenticity of the Other / the Thing (by %)*

Signifies *Authenticity of the Organisation (by %)*

Signifies *Authenticity of the Self (by %)*

Table 3. Validity Check for the Potential New Authenticity Terms (Step 3).
*Terms in bold were new terms identified from the manual selection.

Authenticity Term	Highly Associated Term	Association Score
amateurish	caring	0.642
assuming	pretentious	0.753
authentic	**home-made**	0.632
	house-made	0.543
	ridgy-didge	0.500
caring	**careful**	0.645
classic	traditional	0.685
	typical	0.575
	modern	0.529
	conventional	0.517
crafted	skilful	0.575
	expert	0.563
craftsmanship	**workmanship**	0.715
deceptive	**misleading**	0.619
	deceitful	0.581
exotic	offbeat	0.566
	unique	0.546
fusion	**inspired**	0.635
	modern	0.625
genuine	**heartfelt**	0.616
	fair-dinkum	0.603
heartfelt	sincere	0.674
homey	quirky	0.624
honest	**dinky-di**	0.510
inspired	**fusion**	0.635
misleading	deceptive	0.619
	false	0.569
	honest	0.566
original	**native**	0.522
	true-blue	0.513
sincere	**heartfelt**	0.673
specialty	expert	0.756
	unique	0.654
traditional	**classic**	0.685
	home-style	0.625
	conventional	0.598
	fusion	0.580
typical	**mistakable**	0.598
	classic	0.575
	conventional	0.565
	deadset	0.558
wholesome	**home-made**	0.650
	house-made	0.649

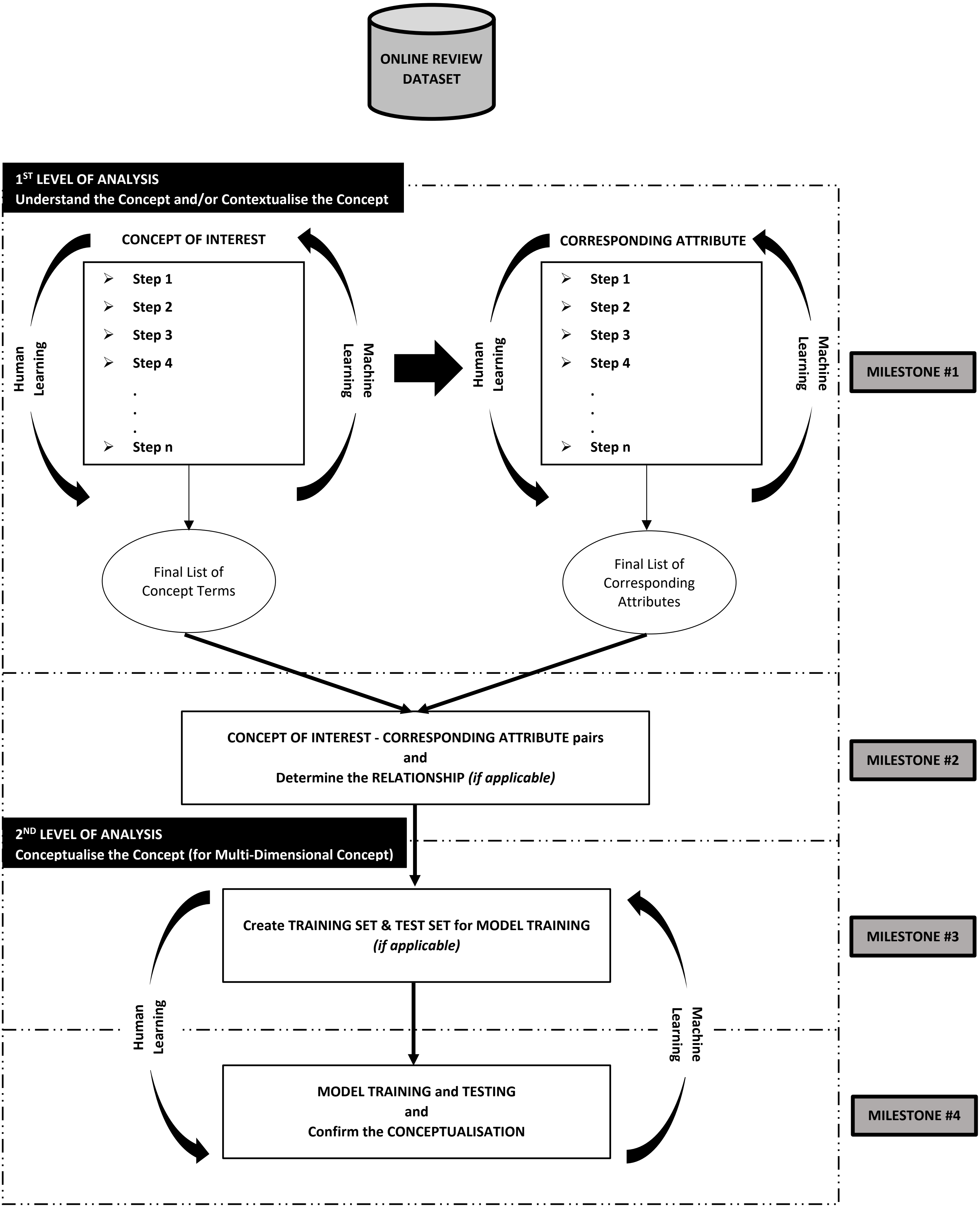

Figure 1. Step-By-Step Methodological and Analytical Framework for the Analysis of Online Reviews.

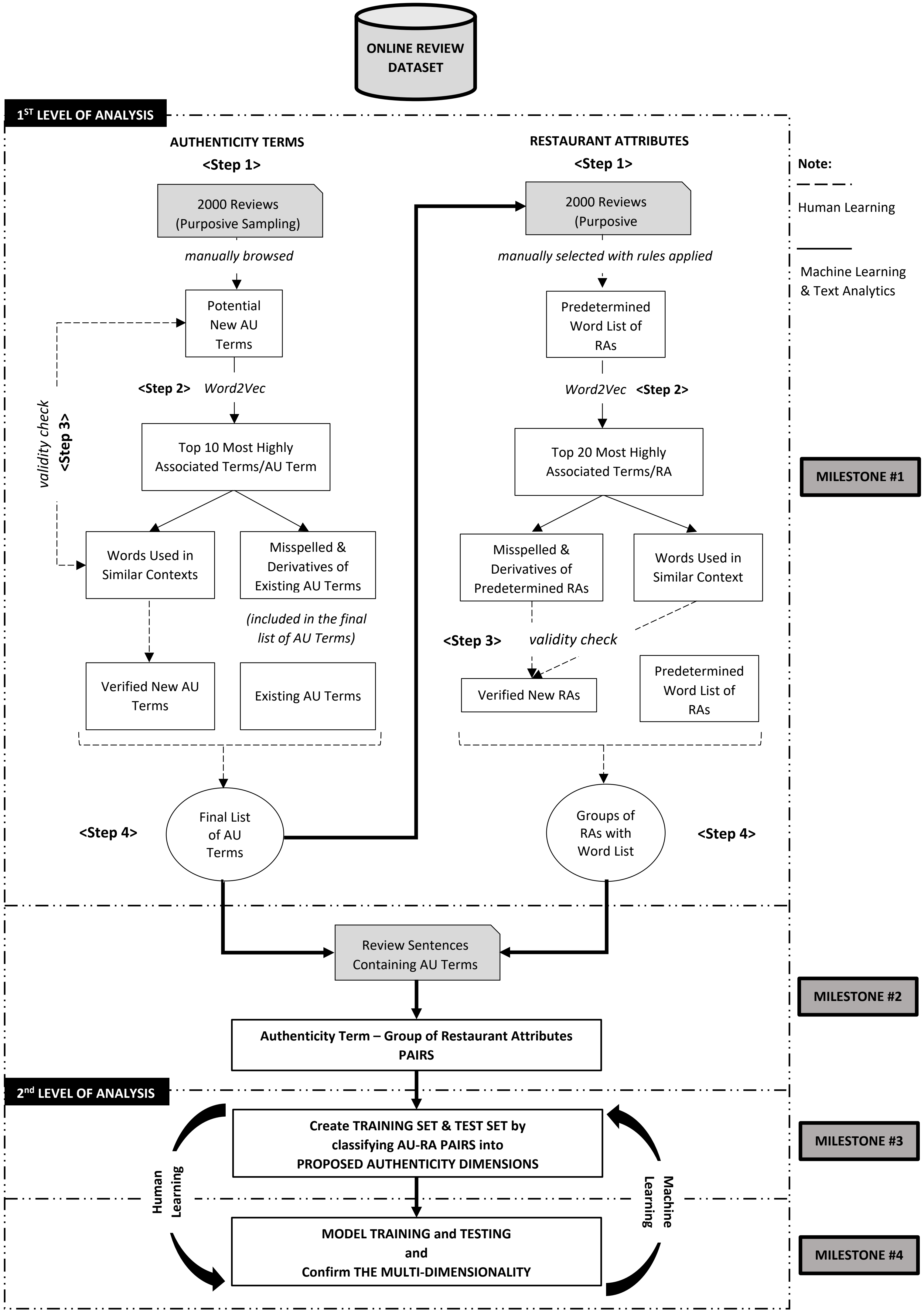

Figure 2. An Example Adapting the Proposed Framework: Authenticity in Dining Experiences.

Figure 3. Final List of Authenticity Terms (Step 4).