Article

Target-Oriented Data Annotation for Emotion and Sentiment Analysis in Tourism Related Social Media Data

Alireza Alaei [1,*], Ying Wang [2], Vinh Bui [1] and Bela Stantic [3]

1   Faculty of Science and Engineering, Gold Coast Campus, Southern Cross University, Gold Coast,
    QLD 4225, Australia
2   School of Hotel and Tourism Management, Hong Kong Polytechnic University, Hong Kong
3   School of Information and Communication Technology, Griffith University, Gold Coast, QLD 4222, Australia
*   Correspondence: ali.alaei@scu.edu.au

Abstract: Social media have been a valuable data source for studying people's opinions, intentions, and behaviours. Such a data source incorporating advanced big data analysis methods, such as machine-operated emotion and sentiment analysis, will open unprecedented opportunities for innovative data-driven destination monitoring and management. However, a big challenge any machine-operated text analysis method faces is the ambiguity of the natural languages, which may cause an expression to have different meanings in different contexts. In this work, we address the ambiguity challenge by proposing a context-aware dictionary-based target-oriented emotion and sentiment analysis method that incorporates inputs from both humans and machines to introduce an alternative approach to measuring emotions and sentiment in limited tourism-related data. The study makes a methodological contribution by creating a target dictionary specifically for tourism sentiment analysis. To demonstrate the performance of the proposed method, a case of target-oriented emotion and sentiment analysis of posts from Twitter for the Gold Coast of Australia as a tourist destination was considered. The results suggest that Twitter data cover a broad range of destination attributes and can be a valuable source for comprehensive monitoring of tourist experiences at a destination.

1. Introduction

Understanding people's opinions, intentions, and behaviours toward a service or product has been critical for various public and private organisations. Traditional research methods for understanding people's opinions and collecting data rely on instruments such as conventional questionnaire-based surveys, interviews, case studies, and focus groups. While these methods offer many advantages, they often suffer from a small study sample size, interviewers' and respondents' bias, and incomplete responses [1].

Social media networks, such as Facebook, Twitter and Web 2.0, were used by locals, tourists, travellers, consumers, and reviewers (hereinafter referred to as users) to virtually share their reviews, comments, photos, videos, emotions, and opinions about anything they experienced in their daily lives. These "electronic word-of-mouth" contents (e-WOM) became a powerful source of information, which is believed to influence consumers' decision-making processes, attitudes, and trust [2,3]. E-WOM are not controlled by product and service providers and their marketing organisations. Therefore, they are considered less biased and more 'trustable' opinions by the public [4]. The power of e-WOM created a new dynamic of influence and persuasion towards various industries, including tourism, given that the tourism industry sells intangible products, which are difficult for consumers to evaluate pre-purchase [5]. As a result, e-WOM became a valuable source of data for studying tourists' sentiments, emotions, opinions, intentions, and

behaviours, opening opportunities for innovative data-driven destination monitoring and management [2,6].

Sentiment in the tourism context is a degree (measure) of feelings (attitudes) towards a tourism destination, service, or product, or an aspect of them, which is embedded in related textual reviews, comments, photos, or videos by tourists and/or marketing organisations [7]. On the other hand, emotion is a human feeling, such as being furious, joyful, angry, cheerful, or sad. Emotion analysis or detection is the way that human emotions are detected from data; for example, text and reviews [8–11]. In the context of tourism, it is beneficial to analyse emotion and sentiment together as such information will give product and service providers and their marketing organisations a good understanding of how tourists perceive their products and services and what areas need further improvements. However, in many previous studies, sentiment and emotion analyses were often carried out independently. This study will address this issue by incorporating emotion and sentiment analysis under a single framework.

Due to the large volume of e-WOM data generated daily [12], there is a growing need for a system that is capable of automatically processing and analysing e-WOM for sentiments, emotions, and other information [5,8,10,13,14]. In the last decade, researchers proposed different techniques to perform sentiment and emotion analysis on big data [7,8,10]. Natural language processing (NLP), text mining, automatic text categorisation, and summarisation are among the techniques used by researchers to automatically process, classify, and summarize e-WOM textual data, such as reviews and comments [12]. More recent studies used lexicon-based and machine learning techniques to measure overall sentiments and emotions associated with e-WOM [6–8,15,16]. Despite the progress in this area, none of the current sentiment analysis techniques were able to perform well across all domains and different contexts. As pointed out by Ribeiro et al. [15], sentiment results given by the existing techniques varied widely even across similar datasets, which indicates that the same content could be interpreted very differently depending on the choice of a sentiment analysis technique and the domain and context of the problem. A study by Kessler et al. [17] also suggested that sentiment polarity when analysing an opinion may be affected by the context in which the opinion was given. For example, the adjective "lazy" has a 'negative' connotation when it is associated with "staff" in the context of "lazy staff".

On the other hand, the "lazy" adjective might be interpreted as 'positive' when someone is "having a lazy afternoon on the beach". Ambiguity is the beauty of a natural language, but at the same time, it poses a difficult challenge for any machine-based sentiment analysis technique. Moreover, sentiment and emotion analysis was mainly carried out at the post level, while a post often contains multiple topics (targets) [7].

A natural approach to overcoming the ambiguity challenge is to limit the context in which sentiment analysis is performed and to develop a context-aware sentiment analysis technique for the context. We, therefore, proposed a method to analyse emotion and sentiment for individual topics/targets under a single framework. We showcase our framework using a limited number of tweets in relation tourism. In particular, we proposed a method for conducting a so-called automatic target-oriented emotion and sentiment analysis of e-WOM. The method is developed to perform sentiment analysis and emotion detection in the context of tourism destination management. The term 'target' refers to a specific point of interest at the destination. Limited data collected from the Gold Coast of Australia as a tourist destination are used to illustrate the proposed system.

We conducted this study in two phases. The first phase aimed to present an annotation approach and construct an annotated dataset for training and assessing the performance of target-oriented emotion and sentiment analysis algorithms. In addition, it also provided (i) a method for annotating social media data for target-oriented and emotion and sentiment analysis of a tourism destination; (ii) a tourism-specific lexicon for target detection; (iii) an extended lexicon for emotion detection, and (iv) annotated sentiment scores for target-oriented sentiment analysis purposes. Although we used Twitter posts to construct the dataset in this study, the method is applicable to other social media data. The dataset

further helped us to create two dictionaries: target and emotion. It was also used to evaluate target-oriented and emotion and sentiment analysis algorithms.

Based on the dataset created in the first phase, the study's second phase aims to develop an automatic target-oriented emotion and sentiment analysis method and assess its performance. In particular, we will analyse the sentiments of individual targets of the destination, the Gold Coast in this case, and compare the results with those obtained by overall sentiment and general sentiment analysis techniques. As the previous works on sentiment analysis of tourism destinations only provided the overall sentiment towards a destination [18], this study will analyse the sentiments and emotions of individual targets of a destination. In summary, our contributions in this research work are (i) proposing a framework for data annotation and creating a dataset of textual data with associated labels, (ii) proposing a method to analyse both emotion and sentiment for an individual topic/target extracted from social media posts, e.g., Twitter posts, and (iii) discussing the findings and results obtained from the annotated data and proposed method to reveal detailed results compared to the previous works.

The rest of the paper is organised as follows. In Section 2, the background of the work is reviewed. Section 3 presents the proposed methodology for creating the dataset and target-oriented sentiment analysis approach. Section 4 discusses the pilot study and results obtained from the proposed methodology. Finally, Section 5 concludes the paper and provides insight for future work.

2. Literature Review

This section provides a brief review of the literature on sentiment analysis and its application in tourism while focusing on target-oriented sentiment analysis and its requirements.

2.1. Sentiment Analysis and Its Application in Tourism

Information provided through a social media post is either objective (i.e., factual) or subjective (i.e., opinionated) [7]. Compared with non-opinionated objective posts on facts, evidence, and measurable observations, subjective posts are the explicit point of view, personal beliefs, feelings and emotions, and judgments about individuals, things or events, which may imply a positive, negative, or mixed feeling [7]. Opinions conveyed in social media posts can be evaluated using sentiment analysis.

Sentiment analysis is a broad field of study, referred to as opinion mining or subjectivity analysis, with some connections to affective computing (computer recognition and expression of emotion) [19,20]. Sentiment analysis is composed of knowledge and technologies from many disciplines, including linguistics, computer science, and natural language processing. In this paper, we refer to sentiment analysis as a computational process aiming to measure and evaluate consumers' and tourists' opinions, feelings, and emotions expressed in social media posts.

The sentiment analysis literature specified quantitative and qualitative measurement schemes to measure sentiments, where the binary, ternary, and ordinal sentiment polarity classification schemes are the most widely used. The binary scheme labels a social media post as either "positive" or "negative". In this scheme, any post is subjective, i.e., it reflects a person's viewpoint and opinion. The ternary scheme tries to differentiate between objective and subjective posts, referring to objective posts as those containing facts, evidence, and measurable observations [7]. Objective posts will be classified as "neutral" in this scheme, while subjective posts are classified as either "positive" or "negative". Unlike the binary and ternary schemes, the ordinal scheme classifies a post using a rating scale or sentiment strength [21].

Methods for analysing the sentiment can be categorized mainly into dictionary-based, machine learning, and hybrid, where dictionary-based methods are commonly used in the tourism sector [15,21–25]. The Valence Aware Dictionary for Sentiment Reasoning (VADER) is one of the dictionary-based methods. VADER was specifically proposed for sentiment analysis of Twitter data [24]. It combines a lexicon/dictionary and a series of

intensifiers, punctuation transformations, and emoticons to compute the sentiment polarity of a post [7,24]. On the other hand, machine learning methods rely on machine learning models to compute sentiments. Support Vector Machines (SVM) and Naive Bayes classifiers [22] are common machine learning models used for sentiment polarity classification. As its name suggests, a hybrid method combines machine learning and dictionary-based technique in one process to obtain the final sentiment polarity of a post [26,27].

Sentiment analysis of e-WOM data gained attraction in the tourism industry as a method to explore tourist experiences. Studies were carried out in different contexts, including airlines [25,28], hotels [29,30], and sport events [31]. Duan et al. [29] showed that sentiment analysis of e-WOM data can give a deeper understanding of what contributes to the overall rating score of a hotel. A similar observation was also reported by Misopoulos et al. [25], suggesting that using e-WOM data can give a deeper insight into tourist experiences beyond what we observed in the literature. Nonetheless, research in this direction remains preliminary, with a small number of works focusing on understanding the overall sentiment of a review or comment. Meanwhile, according to Brob [21], a tourism-related social media post may contain information and opinion on multiple topics and objects of a destination, and the sentiment associated with each topic or object can be analysed individually. This finding suggested that further research can be carried out to develop automatic sentiment analysis methods to target a specific topic, object, service, or product of a tourist destination. As a result, a more sophisticated understanding of a tourist destination's performance can be achieved [7].

2.2. Emotion Analysis and Its Application in Tourism

Emotion classification systems assign labels such as joy, sadness, anger, and fear to sentences. They, too, use feature sets similar to sentiment valence classification systems. In contrast to the sentiment valence classification, there was only one shared task competition (SemEval-2007 Task 14: Affective Text) on detecting emotions organized in 2007 [32], where participants had to determine the emotions in newspaper headlines [10].

Detecting emotions in textual data is sometimes difficult, even for humans. Studies showed that the amount of agreement between annotators is significantly lower in assigning valence or emotions to instances compared to tasks such as identifying parts of speech and detecting named entities. There can be significant differences in emotions associated with events and behaviours across different cultures. For example, dating and alcohol may be perceived as significantly more negative in some parts of the world than in others [10].

Paul Ekman's basic emotions include joy, sadness, anger, fear, disgust, and surprise. Since the writing style and vocabulary in different sources, such as chat messages, blog posts, and newspaper articles, can be very different, automatic systems that cater to specific domains are more accurate when trained on data from the target domain. Holzman and Pottenger [33] annotated 1201 chat messages for Ekman's six emotions as well as for irony and neutral classes. Alm, Roth, and Sproat [34] annotated 22 Grimm fairy tales (1580 sentences) for Ekman's emotions. Strapparava and Mihalcea [32] annotated newspaper headlines with intensity scores for each Ekman emotion, referred to as the Text Affect Dataset. Aman and Szpakowicz [35] annotated blog posts with the Ekman emotions. These datasets were used to develop supervised machine learning algorithms to classify instances into one of the six Ekman classes (or neutral) [10].

2.3. Target-Oriented Sentiment Analysis

Target-oriented sentiment analysis intends to classify sentiment polarities over a target, i.e., a point of interest such as an individual topic, product, service, landmark, or object in an opinion text. A target can be a single entity, such as a product or a service, but it can also be a complex hierarchy of sub-targets [36]. For example, a tourist destination is a target with different aspects (sub-targets), such as attractions, accommodation, transport, and natural environment, which maintain some relationship with each other. In this study, a hybrid annotation approach, combining the top-down coding process informed by the

literature [36–39] on key destination attributes with the bottom-up coding process through which new targets/aspects were added to the literature-informed list of attributes as they emerged in the annotation process were considered for destination structure and nine aspects, including attraction, accommodation, food and beverage, weather, people and culture, transport, holiday in general, shopping and gift, and city facilities, were identified. This is in line with the attributes found by Chen et al. [38].

A challenge in target-oriented sentiment analysis is identifying targets, sub-targets, and their relationships. Based on the granularity of analysis, a target may refer to a concrete tangible entity or an abstract subject that was commented on in a post. Appropriate classification and modelling of targets help to determine more precise target-oriented sentiment. Relations between different aspects of a complex target are often modelled using ontology learning techniques [40]. Depending on the application and the grouping rules, the hierarchical relationship between a target and its aspects or sub-targets may differ. We will discuss the target identification and modelling process in more detail with examples later in the subsequent sections.

To develop machine-based sentiment analysis methods to target specific topics and objects of a tourist destination, relevant datasets with sentiments annotated for the topics and objects of interest are required. Annotated datasets are crucial for training and evaluating the performance of machine-operated algorithms [7]. However, besides a small number of datasets of user-generated reviews (UGR), focusing mainly on domains other than tourism, such as news, movies, meeting minutes and reports [17,36] there are only a few datasets constructed for the hospitality and tourism domain within a small set of areas including hotels [21–23,27], airlines [25] and restaurants [41]. Data sources include Twitter [25,26], Citysearch [41], Booking.com [22], and TripAdvisor [21–23]. More importantly, the analytical focus of these datasets is on the overall sentiment toward a tourism destination, service, or product rather than the sentiment toward a specific attribute of the above. Among the datasets mentioned above, only Brob's [21] dataset of restaurant and hotel reviews annotated targets for target-oriented sentiment analysis. To the best of our knowledge, no publicly accessible dataset is available that comprises annotated social media posts to provide the ground truth of sentiments toward a range of targets of a tourism destination, such as accommodation, food and beverage, attractions, and weather. Subsequently, there is also no existing method or scheme to annotate social media posts for target-oriented sentiment analysis. This study, therefore, proposes a method to annotate social media posts for target-oriented emotion and sentiment analysis and subsequently builds a publicly available annotated dataset from Twitter posts to bridge the research gap.

Sentiment analysis is a fundamental task in natural language processing. Since the sentiments in a sentence can be complex and varied by different targets, target-oriented sentiment analysis is proposed to refine sentiment analysis. Target-based sentiment analysis aims to detect a target and sentiment polarity in a text. Literature studies revealed that a sentiment polarity extracted from a text depends on the target. However, most of the existing sentiment analysis methods predict general sentiment polarities, and in several cases, they may make wrong predictions on sentiment polarities. In particular, where the target is implicit, i.e., it does not appear in the given text, the methods predicting sentiment polarities from targets do not work [42]. Recently target-oriented sentiment analysis gained more and more attention, especially with the rise of social media and public opinion. SemEval-2015 Task 12 [43] and SemEval-2016 Task 5 [44] formalize ABSA as a task for target-aspect-sentiment detection from a sentence. To tackle these limitations, this paper proposes a dictionary-based method for target-oriented sentiment detection. It relies on a target dictionary obtained from a set of target words [42].

Figure 1 gives three review examples in the restaurant dataset from SemEval-2016 Task 5. The first example shows that the sentiment cannot be determined by the sentence and the aspect. The second example shows that the sentiment cannot be determined by the sentence and the target. That is, the sentiment depends on both the target and the aspect.

Moreover, the target of the last example can be given implicitly; i.e., it is assigned NULL and does not contain any word in the sentence [42].

id = 1632445:5 Even though its good seafood, the prices are too high.

target: seafood aspect: FOOD#QUALITY sentiment: positive

target: seafood aspect: FOOD#PRICES sentiment: negative

id = 1189674:3 Admittedly some nights inside the restaurant were rather warm, but the open kitchen is part of the charm.

target: open kitchen aspect: AMBIENCE#GENERAL sentiment: positive

target: restaurant aspect: AMBIENCE#GENERAL sentiment: negative

id = 1726473:4 The food arrived 20 min after I called, cold and soggy.

target: NULL aspect: SERVICE#GENERAL sentiment: negative

target: food aspect: FOOD#QUALITY sentiment: negative

Figure 1. Reviews along with their sentiments, targets, and aspects considered from SemEval-2016 Task 5.

A review of a product or service can express sentiment towards various aspects. For example, a restaurant review can gush positively about the food but express anger towards the quality of service. There is now a growing amount of work in detecting aspects of products and also sentiment towards these aspects [10,17].

3. Methodology

In this section, we present the methodology for collecting and annotating Twitter posts to understand the targets and emotions conveyed in the posts and the methods for target-oriented emotion and sentiment analysis of tourism-related Twitter posts.

3.1. Methodology for Post Collection and Annotation

3.1.1. Annotation Scheme and Key Definitions

The main objective of the annotation scheme is to clearly define the elements of a post and the annotation procedure, so that the sentiment polarity of any target and/or emotion in a post can be assessed manually. We adopted Pang and Lee's [20] proposition to define a post as a hex-tuple (r, t, p, e, h, td), where r is a post, t is the target of the post r, p is the polarity orientation of the post r in relation to the target t, e is the emotion expressed in the post r, h is the post holder of the post r, and td is the time/date when the post r was expressed by h.

Target: in this annotation scheme, a target is defined as a tourism-related entity, which can be found in a post, such as a service, an attraction, an event, an organization, or a person. As a target can be a single or complex entity, we define a target as a pair, t: (T, A), where T is a hierarchical taxonomy of the target and A is a set of aspects/attributes of the target. If the target t is complex, its hierarchical taxonomy will contain sub-targets, i.e., $T = <t_1, t_2, \ldots, t_n>$, n is the number of sub-targets. If the target t has more than one aspect/attribute, then $A = <a_1, a_2, \ldots, a_m>$, m: is the number of aspects/attributes. In our study, the target is tourism, and aspects are the most relevant tourism attributes, such as attraction, accommodation, food and beverage, weather, people and culture, transport, holiday in general, shopping and gift, and city facilities.

In other domain-specific contexts, the annotation team can review the literature related to the domain context to identify targets and their hierarchy or use available ontology databases and semantic search engines for the task.

Opinion passage on a target or aspect: as a post *r* can comprise many sentences, we define the post as $r = <s_1, s_2, \ldots, s_k>$, where s_i is the *i*th sentence in the post and *k* is the number of sentences. If the post r expresses different opinions on multiple aspects/targets, each sentence will be annotated individually in relation to the related target.

Explicit and implicit opinion: a sentence that directly expresses a positive, negative or mixed (neutral) opinion is a subjective sentence with an explicit opinion. An implicit sentence indirectly implies an opinion and can be understood through the context of a whole post/sentence [36]. For implicit opinions, annotators interpret the posts, and the triangulation between the annotators is used to ascertain the target and its sentiment polarity.

Sentiment polarity: this study adopts the ordinal sentiment classification system where polarity is quantified with an interval value to show the intensity of positivity or negativity, with the value 0 representing a neutral polarity. Two annotators manually assign a value between −5 (negative) and +5 (positive) to the polarity of each post as a whole and its meaningful sub-components. The annotators then discussed inconsistencies to reach an agreement. Finally, the average sentiment polarity is computed for each sentence and entire tweet.

Emotion: emotion is a complex mental state which is difficult to measure. Tourism experiences are hedonic in nature, and emotion plays a pivotal role in motivating travel and influencing tourist satisfaction [45]. Conventional self-reported measures of emotion (e.g., emotional scales in a questionnaire) are common in marketing and tourism research. These measures tend to retrospectively assess consumer perceptions of emotions rather than capture emotions in real-time [45]. In this study, we adopted a categorical emotion detection model, which defines a set of six discrete emotional categories, including anger, disgust, fear, joy, sadness and surprise [9]. We also merged emotions of a similar nature into the same category; for instance, the "Anger" category also includes "Disgust", and the "Happiness" category includes emotions such as "Joy" and "Delight" [9,46]. This model allows us to measure emotion in a post, which may explicitly include one or more descriptive words expressing a specific emotion about an aspect of a target. It may indirectly imply an emotion that annotators interpreted in the context of the whole sentence. For example, the post, "Sounds breezy, but looks divine! Heading over to Magnetic Island this morning", implies joy and excitement.

3.1.2. Annotation Procedure

Post collection and annotation is a multi-stage process illustrated in Figure 2. The first part in this process is to collect Twitter posts from the social media platform. The streaming API was used for the continuous collection of tweets for a specific time period. Historically, one per cent of tweets posted on Twitter every day were freely accessible by researchers and they could collect tweets based on a random sampling approach. We used a filtering approach based on locations, where tweets were posted to collect data. This filter with a rectangular bounding box covering the major areas of the Gold Coast helped us to only collect tweets that originated from within the Gold Coast and surrounding suburbs. Uniform random sampling was further used to select a subset of the collected posts for the manual annotation. It is important noting that, in general, people are more positive than negative in their daily lives resulting in a natural bias toward positivity. Therefore, datasets crawled from social media networks, e.g., Twitter, are generally imbalanced. The sample data selected based on uniform random sampling for annotation and further analysis also followed a similar distribution to the crawled dataset. Coordinated by two project team members (one with tourism and the other one with information technology expertise), the annotation process was completed in two steps: training and annotation of the full dataset. Each post was annotated independently by two researchers following the instructions developed by the team of experts (Section 3.1.3), who coordinated the annotation effort and also annotated the training data. A two-stage annotation scheme was proposed and adopted to facilitate the manual annotation process. In the first stage, the selected posts

were initially analysed at the post level to determine the relevance of posts to tourism (i.e., in scope). In the second stage, and for tourism-related posts, a three-level annotation approach consisting of post, sentence, and expression levels was proposed. Annotators were proficient in English and familiar with the subject of tourism.
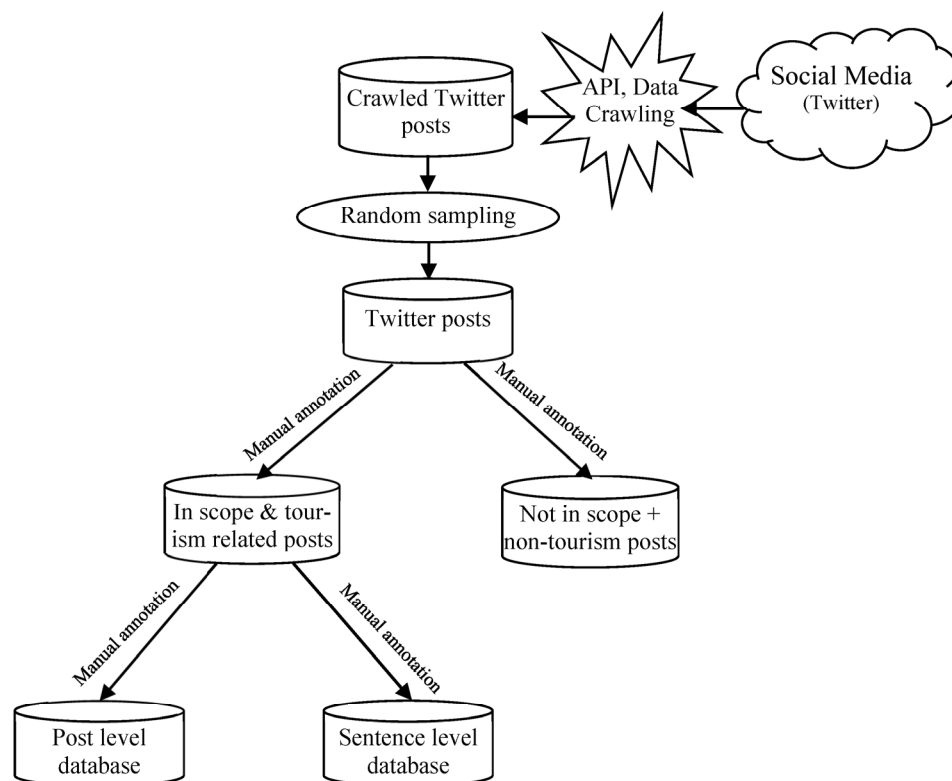

Figure 2. Graphical representation of the procedures applied to create the proposed database in this research work.

　　　Annotation used a hybrid approach, combining the top-down coding process informed by the literature on key destination attributes and emotions with the bottom-up coding process through which new targets/aspects were added to the literature-informed list of attributes and emotions as they emerged in the annotation process. Annotators were provided with the hierarchical system of targets and aspects [36–39] as well as the emotional vocabulary list (e.g., Your emotional vocabulary list by K. McLaren, https://karlamclaren. com [47] (accessed on 12 April 2023) for coding. After reading the annotation instructions, they were trained on 100 Twitter posts, out of which 15 were identified as tourism related. In the training stage, annotators discussed with one another and the team of experts, and they were allowed to ask questions to clarify their understanding of the scheme. Similar training took place in the second stage with the three-level annotation of tourism-related tweets. Training improves the robustness and reliability of annotation. As reported in Section 4.1.1, the inter-annotator agreement is high among annotators. For cases of disagreement between annotators, harmonisation then followed, involving discussion of all four coordinators and annotators to reach one agreed version for each tweet in terms of all properties of the tweet.

3.1.3. Instructions for Coding Tweets

　　　To better understand a text, in general, and a social media post (e.g., a tweet), in particular, the tweet needs to be interpreted in its context. For example, "I can tell you now, it's gonna be GS vs. Ca vs. #yuck #rematch" talks about an NBA match. In this case, we should not be coding into events happening at the destination. Still, it is relevant to the study in the sense that tourists watched/read about a match that happened somewhere

else, affecting their emotions. This point falls into the leisure or recreational aspect of a tourist experience.

Stage 1: Determine in scope and tourism-related reviews

In-Scope: determine whether tweets are codable. Exclude those that cannot be read or analysed, e.g., tweets written in languages other than English, tweets with dominant non-English words, tweets composed of only URLs, and understandable tweets.

Category of Tweeter: code whether a tweet is likely a tourism-related tweet or conveys something about tourism, such as travel, hotel, restaurant, transport, and hospitality.

Level of confidence: indicate your level of confidence for your coding of the category of tweeter on a scale of 0–100 (0 = not confident at all, 100 = full confidence).

Stage 2: Determine polarity, targets, aspects, emoticons, sentences, and words.

Location target: This can be broken down into general, specific, unique, sub-target, and sub-target unique: so, an example of each might be Gold Coast (general), Suburb (specific), Palm Beach (unique), Pool (sub-target), Olympic Pool (sub-target unique)—with the "unique" codes being the name or unique identifier of the actual location—Palm Beach and Olympic Pool—and the others being broader categories of location. Other examples are general: e.g., Gold Coast, Brisbane (e.g., we made a side trip to Brisbane today), specific: e.g., "we toured the city centre" or "a beautiful suburb".

Business target: Similar to location target, this can be broken down into general, specific, unique, sub-target, and sub-target unique: so, an example of each might be recreation, sport, AFL Football, Team, Pies (which is a colloquial term for the AFL football team the Magpies). Location and business/person targets, as often tweets will have both.

Other target: This code is designed for those tweets that do not indicate any location and business targets. However, tweets with location/business target will also be coded again under "other target". We sometimes need to code a tweet into multiple categories of the target. For instance, a Pool can be a location sub-target, a target—general, or a sub-target, depending on the context.

Target-General: the overall target is the natural environment, so you may code it as "natural attraction".

Target-Specific: these are sub-targets. The natural attraction here includes four sub-targets: star (1.1), sky (1.2), sunrise (1.3), and wildlife-bat (1.4).

Aspect: There might be a few aspects in each tweet, so we need to code each separately. An example of this is demonstrated in the following tweet: "5 am so many stars in the sky. Off to the beach to try to catch the sunrise. Giant bats flying around."—In this case, no location or business is mentioned, but there are still some aspects.

Aspect: aesthetic. In this case, the tweet discusses the aesthetic aspect of all four sub-targets. It is also possible that a tweet discusses different aspects. E.g., "room is clean, staff is friendly". Aspects would be cleanliness for the room and friendliness of the staff.

Descriptors—giant, flying (for wildlife), so many (for stars) are often adjectives.

Another example is the post "Even though it is a good seafood restaurant, the prices are too high." This post contains two parts, of which the first part is composed of "seafood restaurant" as the target and "food quality" as an aspect. The second part also talks about a "seafood restaurant" as the target and "food prices" as an aspect.

The intensity of polarity: Polarity relates very closely to valence and emotion research. Generally, the linguistics and emotion literature suggest coding valence as positive, negative, neutral, mixed, or equivocal. However, the emerging "Big Data" literature coded it simply as positive, negative, and neutral. For an overall tweet, we are not considering mixed and equivocal valence; we hard code them into a scale from −5 to 5. Code where you think the tweet would approximately sit on this 11-point intensity scale. The "5 pm so many stars . . . " example above is an example of a relatively neutral tweet, as it describes things without much positive/negative emotions. Sometimes, a tweet comprises more than one valence/emotion and can be a mix of positive and negative feelings. For example, "@WannabeKimba Awww sorry to hear you had a bad night. We all have them, but things

do get better! Go do something u enjoy so u feel better!". In this case, we need to code first the polarity for the whole tweet.

Second, *break the tweet down to its meaningful components/sentences*. A component here is defined as a textual unit consisting of one or more grammatically linked words to convey a meaning (in big data, known as "sentence").

Then, for each component/sentence, we code the intensity again on a scale ranging from −5 to 5.

Hedging and superlatives: identify hedging and superlatives which indicate the use of linguistic politeness or promotional techniques.

Emotional words and their intensity: identify emotional words and code their intensity *individually*. Emotion lexicon files [48] were used to generate an idea of emotions and their intensity levels in each group. When the vocabulary words are unavailable, code based on where you think the tweet would approximately sit on the 5-point intensity scale, and then code the emotional words and the feelings they are having so that we can empirically link them later.

The emotion they are feeling: Describe in one or two words a tweet's feelings, e.g., frustration, excitement, happiness, fear, etc.

Linguistic markers (Emoticons): We also code separately the linguistic markers (e.g., :-), :), ☺, :(, :-(, ☹, !!!, ???, hahaha, lol) as these provide insight into the intensity and emotion, as well as when the vocabulary is not matching the markers (instances of sarcasm).

3.1.4. Post and Sentence Level Annotation

Post-level annotation examined each post in terms of (i) its relevance to a target and (ii) its polarity in relation to the target determined in (i). Emotion orientation and emotion words were also identified in this stage. It is worth noting that manual annotation was able to deal with situations such as (1) negations (e.g., 'but', 'however' and other words that deny a previous condition) and modifiers (e.g., capitalisation of letters, multiple punctuations to emphasise a point) that alter the polarity of an expression of sentiment; (2) sentiment of a target that is affected by other targets; and (3) other less understood phenomena, such as sarcasm and tone, which are also important to accurately determine the polarity of a post or a sentence [17]. Following the identification of targets of the posts, sentence-level annotation determined the velocity of the polarity of each sentence. Then, word-level annotation identified and categorised emotional words in a post into major categories of emotions.

As a result of manual annotation, various target words were collected and included in a target dictionary. Two experts then monitored the target words, and similar words were eliminated from the list. Finally, a dictionary of 1932 target words was obtained at the end of this process. Each word may belong to more than one target. Similarly, the words identified in relation to emotions and feeling from posts during the annotation process were collected as an emotion dictionary. We further included classified/identified emotion words used in the literature [9,46] to finally obtain a dictionary of 9390 emotion words.

3.2. Proposed Target-Oriented Emotion and Sentiment Analysis

We propose to use a dictionary-based approach to perform target-oriented emotion and sentiment analysis in this study. The proposed approach is illustrated in Figure 3. This system is able to automatically classify and perform sentiment analysis of travellers' Twitter posts in relation to targets and emotions concerning a tourist destination. In the proposed system, posts are initially crawled from Twitter and saved in a JSON format database. Only text content from a tweet is used for the analysis. In the pre-processing step, each tweet is cleaned by removing white spaces, hyperlinks, hashtag signs, and @names, and then converted to lowercase characters. The cleaned tweet is segmented into words, and stop words are removed from the list of words extracted from each tweet.
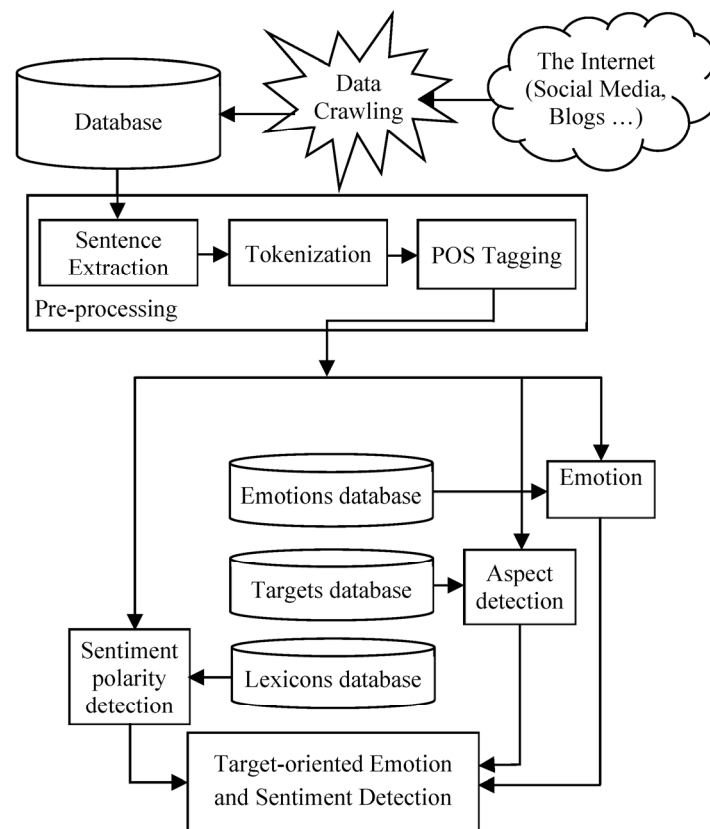
Figure 3. Block diagram of the proposed target-oriented sentiment detection system.

Three different dictionaries, lexicon polarities, targets, and emotion words, are considered. For lexicon polarities, the proposed system uses VADER [24] dictionary, as one of the best lexicon dictionaries for sentiment analysis in the literature [15,49]. The VADER dictionary comprises more than 7000 lexicons along with their associated sentiment intensity measures, which are specifically adapted to sentiment in microblog-like contexts, such as Twitter. The VADER sentiment analysis [24] uses five general rules, which embody grammatical and syntactical conventions for expressing and emphasising sentiment intensity, to compute the sentiment polarity.

As mentioned before, the manual annotation work allowed us to create a target dictionary composed of 1932 words (each word may belong to more than one target) for target detection purposes. These targets were grouped into nine different targets: attraction, food and beverage, accommodation, holiday in general, people and culture, transport, weather, shopping and gift, and city facilities. The proposed system considered these nine key targets identified in the annotation step for target-oriented sentiment detection. Since in the pre-processing stage, a stemming procedure was applied to the extracted words of a post, the stemming process was also applied to the words in the target dictionary, and the stemmed target words were also included in the target words.

A context-aware wordcount normalisation-based approach was proposed to assign a target label to each post. It is called a context-aware approach, as we compute this value considering the corresponding target dictionary. This method further provides a normalised wordcount value between 0 and 1 for each target to indicate the belongingness of a post to that target. To do so, let M be the number of words obtained from a post (R) after pre-processing. The likelihood of a tweet with M words that belongs to a class target $T_j \left(D_{T_j} \right)$ is:

$$P\left(R_{T_j}\right) = \sum_{i=1}^{M} p(w_R^{ij}) \qquad ; \qquad 0 \leq P\left(R_{T_j}\right) \leq 1$$

Each word from the post R (w_R^{ij}) can contribute to this context-aware normalised wordcount value, $P\left(R_{T_j}\right)$, using the following equation:

$$p(w_R^{ij}) = \begin{cases} \frac{1}{M}, & if\ w_R \in D_{T_j} \\ 0, & otherwise \end{cases}$$

where $p(w_R^{ij})$ is the normalised value of the word (i) from the post R belongs to the class target T_j and D_{T_j} is the dictionary of the target class T_j. As we have nine target dictionaries in our proposed target detection approach, j varies from 1 to 9, representing nine different targets. The class target with the highest normalised wordcount value is considered the target of the given tweet. For example, consider "Can't wait for tomorrow. Tambourine mountains and glow worm caves and a long ass hike" as the review R. Applying the preprocessing step, we will have "wait tomorrow tambourin mountain glow worm cave long ass hike" as the pre-processed review R. Now, by applying the above two equations, and considering the nine target dictionaries, the context-aware normalised wordcount values for the attraction (1), accommodation (2), food and beverage (3), weather (4), people and culture (5), transport (6), holiday in general (7), shopping and gift (8), and city facilities (9), the following values were computed. Considering the "Attraction" dictionary, the context-aware normalised wordcount that defines the belongingness to this class is $P\left(R_{T_1}\right) = \sum_{i=1}^{10} p(w_R^i) = 0 + 0 + 0.1 + 0.1 + 0 + 0 + 0.1 + 0 + 0 + 0.1 = 0.4$, as tambourin, mountain, cave, and hike are in the dictionary of attraction. The rest $P\left(R_{T_2}\right) = \sum_{i=1}^{10} p(w_R^i) = 0 + 0 + 0 + 0 + 0 + 0 + 0.1 + 0 + 0 + 0 = 0.1$, $P\left(R_{T_3}\right) = 0.1$, $P\left(R_{T_4}\right) = 0$, $P\left(R_{T_5}\right) = 0$, $P\left(R_{T_6}\right) = 0.1$, $P\left(R_{T_7}\right) = 0.2$, $P\left(R_{T_8}\right) = 0$, $P\left(R_{T_9}\right) = 0.1$ are the normalised wordcount values of other targets. Since the highest value belongs to class (1), the target of this tweet will be considered Attraction (1). It is worth noting that in the case of a tie in finding the highest value of normalised wordcount, we chose the more frequent class, as the predicted target class in the proposed system. For example, if there was a tie between "Attraction" and "Accommodation" normalised wordcount values, since reviews with attractions appeared more frequently in the dataset compared to accommodation-oriented reviews, the "Attraction" class was assigned to the given review.

For emotion detection, we employed the same approach proposed for target detection with only one difference in the number of emotion classes. The emotion dictionary comprises six categories: Happy, Excitement, Love, Fear, Anger, and Sad. A similar method to target detection was further used for context-aware normalised wordcount—based emotion detection to compute a normalised wordcount value between 0 and 1 for each emotion to indicate the belongingness of a post to that emotion. For example, using the same equations and emotion dictionaries, the review "yay for Taronga Zoo! had a blast today" is classified as the "Happy" emotion and the "Attraction" target.

4. A Pilot Study

This section presents a case study of the Gold Coast, Australia. The case study is conducted to reveal some insights and findings considering the proposed annotation approach.

4.1. Data Collection

We collected 102,170 Twitter posts using our API. The posts were primarily written in English and exhibited a rather informal style, often with grammatical errors, misspellings, slang words, and emoticons. Manual annotation requires a high level of human resource input, and so, we limited the database size to a subset of 6000 posts randomly sampled from the database of 102,170 posts. Annotators initially screened the 6000 posts to remove ineligible posts (e.g., websites only and non-English posts). The remaining posts were then annotated into two sets: tourism-related and non-tourism-related. This filtering process could be performed automatically. However, manual screening was employed to make

sure only tourism related reviews were selected. As a result, 475 tourism-related posts were identified for the subsequent sentence and post-level annotations (Table 1).

Table 1. Descriptive summary of the annotated tourism database.

Characteristic	Number
Total number of posts	475
Positive posts	279
Negative posts	28
Neutral posts	168
Total number of sentences	585
Positive sentences	342
Negative sentences	43
Neutral sentences	200
Tweets composed of 1 sentence	382
Tweets composed of 2 sentences	76
Tweets composed of 3 sentences	17

As shown in Table 1, most of the posts (279) were positive, and a small proportion (28 posts) was negative. The total number of sentences obtained during the annotation process was 585 sentences.

4.1.1. Annotation Agreement

Cohen's kappa coefficient (κ), a measure of inter-rater agreement of qualitative items [50], was used to assess the agreement between annotators. This measure is more robust than the simple per cent agreement measure, as it accounts for the possibility of the agreement occurring by chance [51]. The test results suggest a relatively high level of agreement between the two annotators for in-scope posts ($\kappa = 0.877$, $p < 0.0005$) and tourism-related posts ($\kappa = 0.868$, $p < 0.0005$). We also computed the agreement between annotators regarding the sentiment of the overall posts, their components/sentences, and identified emotional words, as measured by the correlation coefficient. The agreement between annotators at the post level was 74.9%, increasing to 79.4% at the sentence level and 92.3% for emotional words, providing a reasonable basis for subsequent analyses.

4.2. Discussion on the Annotated Data

Tables 2 and 3 are the descriptive summaries of our proposed annotation procedure. Confidence intervals (CI) with a 95% confidence level were also computed for the total number of tweets with different targets and emotions. To better understand the summaries, the pictorial presentations of the results are also provided in Figures 4–6, respectively. Table 2 provides the results of the annotation process related to key targets and emotions. The far-right column presents the total numbers and percentages (with CI) of posts for ten targets (i.e., the nine key targets identified in annotation and an "other" category). The bottom row displays the total numbers and percentages (with CI) of posts conveying each emotion category. The "other" category includes targets, such as sports and events. The top three most talked about targets were attractions (manmade or natural), accommodation, and food and beverage, highlighting that these targets matter most to tourists and form the 'backbone' of their trip. The dominance of attraction-related posts reflects their critical role as a pull factor of a destination. In contrast, shopping and city facilities attracted little coverage, confirming earlier work by Volo [52], who did not reveal these as recurring themes in bloggers' reported tourist experiences. From Table 2, it can further be noted that there were a total of 567 instances of positive emotions, 237 instances of negative emotions, such as sadness, anger, or fear, and 141 posts without expressing any emotions. This result confirmed that people tend to share positive experiences [21,53], but the extent to which this distribution resulted from positive bias among Twitter users requires further investigation.

Table 2. Targets vs. emotions detected from tweets based on the proposed annotation scheme.

	Emotion							
	Happiness	Love	Excitement	Sadness	Anger	Fear	No Emotion	Total No of Tweets (95% CI)
Attraction	206	117	95	61	44	52	68	311 (65.47 ± 4.28%)
Accommodation	65	47	27	33	20	21	16	113 (23.79 ± 3.60%)
Food & Beverage	55	43	30	24	22	18	6	95 (20.00 ± 3.23%)
Weather	66	45	43	22	16	18	11	94 (19.79 ± 2.92%)
People & Culture	59	43	32	28	27	22	2	72 (15.16 ± 1.57%)
Transport	43	29	20	18	11	14	13	67 (14.11 ± 4.28%)
Holiday in general	43	23	23	18	14	11	5	57 (12.00 ± 3.83%)
Shopping & Gift	12	13	8	4	3	2	5	23 (4.84 ± 3.60%)
City Facilities	10	9	12	4	4	3	1	15 (3.16 ± 3.23%)
Other targets	12	6	4	3	2	3	38	51 (10.74 ± 2.92%)
Total No of tweets (95% CI)	278 (58.53 ± 4.43%)	163 (34.32 ± 4.27%)	126 (26.53 ± 3.97%)	96 (20.21 ± 3.61%)	70 (14.74 ± 3.19%)	71 (14.95 ± 3.21%)	141 (29.68 ± 4.11%)	

Note: Total number of tweets adds up to more than 100% because some posts belong to multiple target/emotion categories. For instance, a post on a restaurant in a hotel relates to both the food and beverage and the accommodation categories. Similarly, the analysis identified more than one emotion from some Twitter posts.

Table 3. Sentiment polarity of targets detected from the tweets based on the proposed annotation scheme.

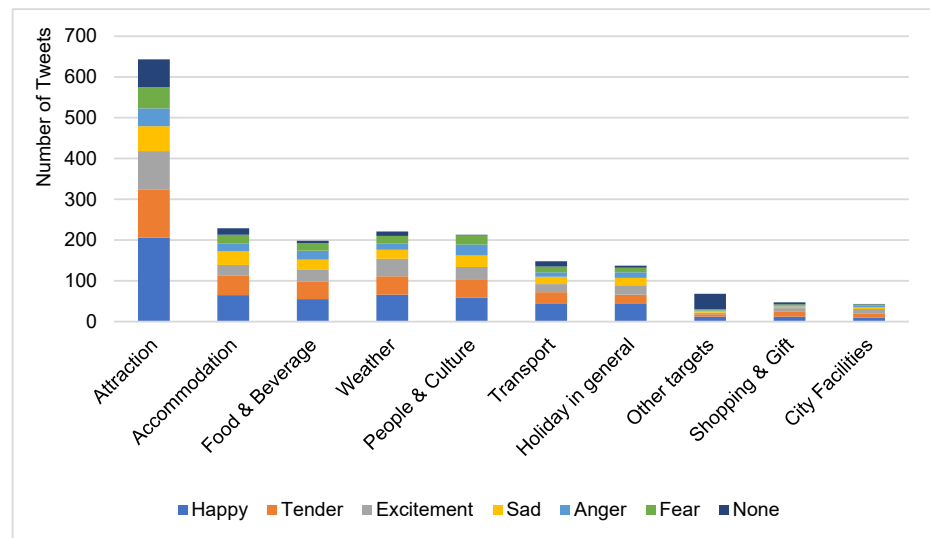| Target	Number			
	Positive	Negative	Neutral	Total
Attraction	144	13	154	311
Food & Beverage	40	8	47	95
Accommodation	47	7	59	113
Holiday in general	29	3	25	57
People & Culture	43	9	20	72
Transport	23	9	35	67
Weather	43	3	48	94
Shopping & Gift	11	2	10	23
City Facilities	7	2	6	15
Other	28	1	22	51
Overall (95% CI)	279 (58.73% ± 4.43%)	28 (5.89% ± 2.12%)	168 (35.37% ± 4.30%)	

Figure 4. Targets and emotions detected from tweets based on the proposed annotation scheme.
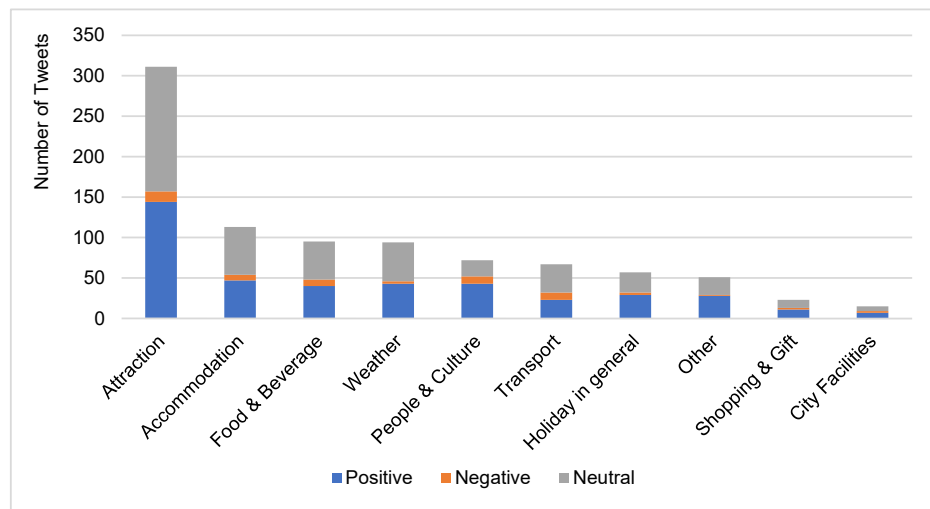

Figure 5. Sentiment of the targets detected from the tweets based on the proposed annotation scheme.
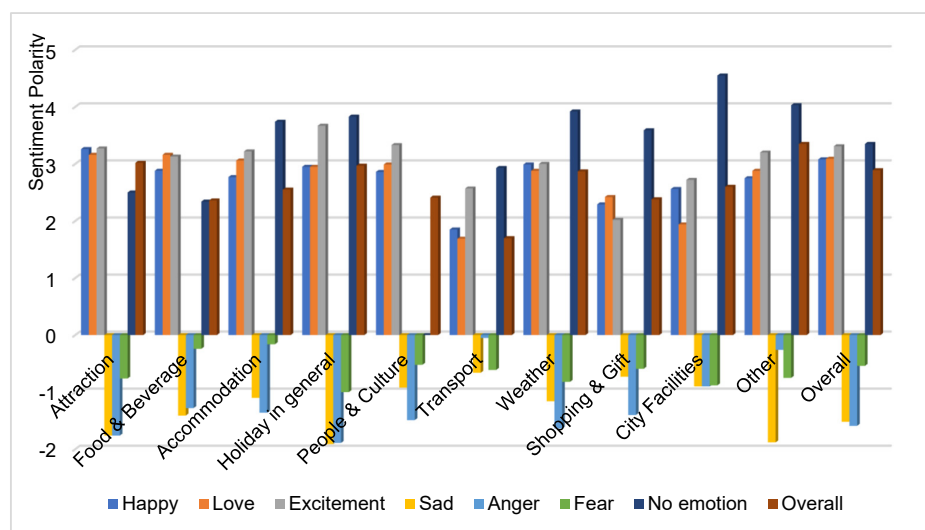

Figure 6. Sentiment of the targets and emotions detected from the tweets based on the proposed annotation scheme. Note: the automatic system generated results in a range of −1 to +1. We converted the results to the range of −5 and +5 to align with the results of our manual annotation.

From Figure 4, it is evident that targets were associated with certain types of emotion. For example, a large number of posts classified as attraction target were associated with happiness. The same observation can also be seen from other targets. This means people generally expressed and shared their happiness with others on our data.

An analysis of sentiment polarity cross-referenced to targets is also provided from the annotated data in Table 3. Table 3 shows that around 35% (±4.3%) of the posts with a 95% confidence level contained neutral sentiments and the rest were either positive (58.73% ± 4.43%) or negative (5.89% ± 2.12%). Table 3 further indicates that the attraction target presented the highest share of posts across the three sentiment polarity categories. The posts with Accommodation, Food & Beverage, and Weather, as the other three targets, were also posted and shared by people in the data considered for this study. Figure 5 also indicates that posts with positive sentiment were more than the posts with negative sentiment. When interpreting the results, it is important to consider the nature and impacts of the low number of tweets in some findings and discussion around them. Particularly, findings and discussion about negative tweets in our dataset should be interpreted with precautions, as negative tweets were small in number and the negative sentiment about the "Holiday in general" target, for example, was computed based on three tweets.

Figure 6 presents the sentiment scores of annotators for targets by emotions. Overall, posts were most positive about attraction holiday in general and the "other" category. The "other" category includes targets, such as sports and events, which generally include a higher level of personal involvement, resulting in a high sentiment score. Transport, food and beverage, and shopping and gift indicated less positive sentiment, suggesting that these are the least performing aspects of the destination.

Our annotation process further revealed that posts were more positive when they did not contain any emotional words. This may sound odd, but positive words (e.g., "good", "right") often do not convey emotions. There were also a small number of tweets containing words/phrases, such as "Surfers Paradise" (name of the most popular tourist beach) and "Dreamworld" (name of a popular theme park). Emotion-wise, posts were most positive in relation to emotions of excitement, love, and happiness. Sentiment polarities were most positive when posts expressed excitement about holidays in general, people and culture and attraction, as well as happiness about attraction.

On the contrary, sadness, anger, and fear generally imply negative sentiments. Considering the data annotated in this study, posts were mainly negative when sadness and anger were shown in the posts.

4.3. Performance of the Proposed Sentiment Analysis System

Accuracy (A), precision, recall, and F-measure are commonly adopted metrics for evaluating the performances of sentiment analysis [7,15,21]. The performance of the proposed sentiment analysis system was evaluated using the F-measure, and the results are shown in Tables 4–6. It is worth mentioning that the proposed system was built based on the VADER and target and emotion dictionaries. The VADER [24] without any further training considering its original parameters fine-tuned during its construction was used as the sentiment extraction in our proposed system. In addition, as there were no parameters in our proposed target and emotion detection steps, there was no need for training the system during the evaluation process. Therefore, the entire dataset was considered in our experiments for testing the proposed method [54].

Table 4. Comparison of the F-measures (FM) computed at the post level results (emotions) obtained from the proposed system and NRC emotion lexicon.

Emotion	Sentiment Polarity (%)							
	Happy	Love	Excitement	Sad	Anger	Fear	None	Average
NRC Emotion Lexicon	52.85	54.55	46.15	56.25	41.10	43.75	**80.74**	64.84
Overall	**70.04**	**69.58**	**68.87**	**59.52**	**58.95**	**63.77**	64.83	**65.59**

Table 5. F-measures (FM) computed at the post level results (targets) obtained from the proposed system.

Target	Sentiment Polarity (%)
Attraction	66.19
Food & Beverage	65.39
Accommodation	66.36
Holiday in general	64.96
People & Culture	66.07
Transport	65.90
Weather	66.13
Shopping & Gift	62.96
City Facilities	65.17
None	66.84
Overall	65.59

Table 6. Performance of the proposed system in terms of sentiment analysis at sentence and post levels.

Metric Method	Neutral			Positive			Negative			Overall Accuracy (%)	Correlation (Proposed System vs. Human)
	Precision (%)	Recall (%)	F-Measure (%)	Precision (%)	Recall (%)	F-Measure (%)	Precision (%)	Recall (%)	F-Measure (%)		
Sentence Level	53.8	89.5	67.2	90.0	55.3	68.5	42.9	41.9	42.4	66.0	0.631
Post Level	57.0	89.3	69.6	89.0	57.7	70.0	32.3	35.7	33.9	67.6	0.625

Tables 4 and 5 present the results of emotion and target detection combined with sentiment analysis resulting in target-oriented sentiment analysis. The results obtained from our proposed system are shown in Table 4, indicating that the proposed system performed better on tweets with positive emotions (e.g., Happy and Love) compared to negative ones. In addition, it is also evident that the proposed system performed better on detecting emotions compared to the National Research Council Canada (NRC) emotion lexicon-based method [54]. NRC emotion lexicon-based method [54] provided better results on tweets without any emotions (None). This is because the NRC method [54] normally provided probability-based results for all emotions and when the probability value for all emotions was zero, this indicated that no emotion was detected for the given tweet. Likewise, Table 5 shows that accommodation, attraction, and weather have higher F-measures, indicating better correlations with manual annotation results. Overall, the proposed system can provide target-oriented sentiment analysis, offering more insight into different aspects of tourism and visitor behaviour.

To evaluate the performance of human (manual annotation) vs. machine (our proposed system), we computed the results based on different metrics, including correlation, accuracy (A), precision, recall, and F-measure, as shown in Table 6. From the results shown in Table 6, it is evident that the overall accuracy of the sentiment analysis results at the post level (67.6%) is slightly better than that at the sentence level, with 66% correct prediction. In general, the system performed better on positive and neutral posts than on negative posts.

This might be because of the larger number of positive tweets compared to the negative numbers that make the dataset imbalanced. The correlation between human-annotated results and the results obtained by the proposed machine-operated method at the sentence level is similar to that at the post level. It is worth noting that we further used the sentiment aware nearest neighbour (SANN) model [55] for further experiments. The results obtained from the SANN model were between 3.7% and 13.3% lower than those obtained from the proposed model in this research work.

5. Conclusions and Future Work

Our study contributes to tourism sentiment analysis in several ways. First, it addresses the lack of target-oriented databases for tourism by creating an annotated database grounded on destination attributes (i.e., targets). An annotation scheme was established following a series of training and testing for applicability in tourism. The scheme can relatively easily be adapted to other domains and as such constitutes a valuable resource for other applications of opinion mining. The second contribution relates to identifying destination aspects and emotions that people tend to share in Twitter posts, which subsequently led to a target dictionary explicitly created for tourism sentiment analysis. Our analysis suggests that Twitter data can cover a broad range of destination attributes, such as overall holiday experience, weather and culture, providing consistent results with the findings reported in the literature.

Further, our study demonstrates a novel approach to measuring real-time emotions in tourism. Sentiment scores of emotions by targets reveal interesting insights into the specific sources of emotional expressions. For instance, people expressed stronger happiness, love, and excitement with attractions. In other words, attractions are likely to trigger stronger emotional reactions. Posts expressed a strong sense of excitement with regard to the holiday in general, people and culture, and attractions, but less so in relation to shopping and gift, transport, city facilities, and the weather. Identifying targets and emotions further enabled the development of applications featuring target extraction, polarity assessment, and target-oriented sentiment analysis. The proposed target-oriented sentiment analysis method is a machine-operated system that can effectively analyse tweets to detect targets, identify emotions, and extract sentiment polarity, offering an alternative solution to monitoring destination performance. It is also noted that domain specific dictionary-based methods are effective in a particular domain. However, their performance is heavily dependent on the dictionary and may not perform well in other contexts.

In addition, the proposed annotation procedure brought helpful insight into different aspects of tourism and visitor behaviour. Our analysis suggests that despite the restriction on word count, short Twitter posts cover a broad range of destination attributes, from overall holiday experience to specific aspects of a destination, to intangible destination elements, such as weather and culture. The results are generally in line with those obtained from travel blogs and social media reviews [8,10,52], indicating consistency in findings but the ability to scale up using big data approaches.

Practically, the findings can also assist destination managers with monitoring visitor satisfaction, destination image, and performance. For example, sentiment analysis can assist with areas that require further attention on the part of destination managers. It is important to note that the complex nature of human emotions complicates the analysis of sentiment and tourist experience and is an area requiring further research attention using new technologies. For example, ChatGPT, as a significant development for text generation, is a helpful tool for text understanding and extracting insights in relation to a general topic. However, domain specific dictionary would be helpful to extract explicit insights for a specific topic.

Future Research Directions

The study had limitations, which point to potential avenues for future research. First, our results were based on Twitter data on one city with a particular destination image: a fun

beach destination. Second, manual annotation was performed on only 6000 Twitter posts, while this can be considered adequate for the annotation process, albeit a relatively small sample size, for the performance evaluation of a system. This may also impact the generalisation of some of the findings. Third, the information conveyed through tweets may differ from that carried out in other types of e-WOM, such as travel blogs and review sites.

Moreover, the most commonly expressed emotion in the database was found to be happiness. Whilst it is possible that regardless of destination type, it is human nature to express happiness and joy more often than other types of emotions, it is also possible that the feelings expressed in Twitter posts may be skewed towards particular types of emotions. Another issue identified by this study is the possibility of influencing the sentiment scores by the presence of positive words (e.g., names of locations and tourist attractions such as 'Great') in reviews, especially when the reviews were of short length. Future research shall focus on further improving the performance of the proposed system, which can be achieved by (1) proposing a hybrid approach considering machine learning and dictionary-based approach, (2) exploring weighted context-aware normalised wordcount in target and emotion detection, (3) extending the testing to other destination types, (4) incorporating a wider range of destination attributes as targets in the analysis, (5) extending to more implicit tourism social media data, such as reviews on TripAdvisor and travel/hotel booking sites, and (6) exploring the use of ChatGPT for review/text understanding in the domain of tourism and hospitality.

References

1. Wang, Y. More Important than ever: Measuring Tourist Satisfaction. Griffith Institute for Tourism Research Report No 10. 2016. Available online: https://www.griffith.edu.au/__data/assets/pdf_file/0029/18884/Measuring-Tourist-Satisfaction.pdf (accessed on 12 April 2023).
2. Ladhari, R.; Michaud, M. eWOM effects on hotel booking intentions, attitudes, trust, and website perceptions. *Int. J. Hosp. Manag.* **2015**, *46*, 36–45. [CrossRef]
3. Nieto-García, M.; Muñoz-Gallego, P.A.; González, Ó. Tourists' willingness to pay for an accommodation: The effect of eWOM and internal reference price. *Int. J. Hosp. Manag.* **2017**, *62*, 67–77. [CrossRef]
4. Chiu, C.; Chiu, N.-H.; Sung, R.-J.; Hsieh, P.-Y. Opinion mining of hotel customer-generated contents in Chinese weblogs. *Curr. Issues Tour.* **2015**, *18*, 477–495. [CrossRef]
5. Confente, I. Twenty-five years of word-of-mouth studies: A critical review of tourism research. *Int. J. Tour. Res.* **2015**, *17*, 613–624. [CrossRef]
6. Becken, S.; Alaei, A.; Wang, Y. Benefits and pitfalls of using tweets to assess destination sentiment. *J. Hosp. Tour. Technol.* **2019**, *11*, 19–34. [CrossRef]
7. Alaei, A.R.; Becken, S.; Stantic, B. Sentiment Analysis in Tourism: Capitalizing on Big Data. *J. Travel Res.* **2019**, *58*, 175–191. [CrossRef]
8. Acheampong, F.A.; Wenyu, C.; Nunoo-Mensah, H. Text-based emotion detection: Advances, challenges, and opportunities. *Eng. Rep.* **2020**, *2*, e12189. [CrossRef]
9. Ekman, P. An argument for basic emotions. *Cogn. Emot.* **1992**, *6*, 169–200. [CrossRef]
10. Mohammad, S.M. Sentiment analysis: Automatically detecting valence, emotions, and other affectual states from text. In *Emotion Measurement*; Woodhead Publishing: Cambridge, UK, 2021; pp. 323–379.
11. Wolny, W. Emotion Analysis of Twitter Data That Use Emoticons and Emoji Ideograms. In Proceedings of the 25th International Conference on Information Systems Development (ISD2016 POLAND), Katowice, Poland, 24–26 August 2016; pp. 476–483. Available online: https://aisel.aisnet.org/isd2014/proceedings2016/CreativitySupport/5/ (accessed on 12 April 2023).
12. Xiang, Z.; Schwartz, Z.; Gerdes Jr, J.H.; Uysal, M. What can big data and text analytics tell us about hotel guest experience and satisfaction? *Int. J. Hosp. Manag.* **2015**, *44*, 120–130. [CrossRef]

13. Liu, C.-C.; Yang, T.-H.; Hsieh, C.-T.; Soo, V.-W. Towards Text-based Emotion Detection: A Survey and Possible Improvements. In Proceedings of the International Conference on Information Management and Engineering, Kuala Lumpur, Malaysia, 3–5 April 2009; pp. 70–74.

14. Toprak CJakob, N.; Gurevych, I. Sentence and expression level annotation of opinions in user-generated discourse. In Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, ACL'10, Uppsala, Sweden, 13 July 2010; pp. 575–584.

15. Ribeiro, F.N.; Araujo, M.; Goncalves, P.; Goncalves, M.A.; Benevenuto, F. A Benchmark Comparison of State-of-the-Practice Sentiment Analysis Methods. *EPJ Data Sci.* **2016**, 23. Available online: https://epjdatascience.springeropen.com/articles/10.1140/epjds/s13688-016-0085-1 (accessed on 12 April 2023).

16. Rossetti, M.; Stella, F.; Zanker, M. Analyzing user reviews in tourism with topic models. *Inf. Technol. Tour.* **2016**, *16*, 5–21. [CrossRef]

17. Kessler, J.S.; Eckert, M.; Clark, L.; Nicolov, N. The 2010 ICWSM JDPA sentiment corpus for the automotive domain. In Proceedings of the 4th International AAAI Conference on Weblogs and Social Media Data Workshop Challenge, Washington, DC, USA, 23–26 May 2010.

18. Tubishat, M.; Idris, N.; Abushariah, M.A.M. Implicit aspect extraction in sentiment analysis: Review, taxonomy, oppportunities, and open challenges. *Inf. Process. Manag.* **2018**, *54*, 545–563. [CrossRef]

19. Pang, B.; Lee, L. A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics, Association for Computational Linguistics, Barcelona, Spain, 21–26 July 2004; pp. 271–278.

20. Pang, B.; Lee, L. Opinion Mining and Sentiment Analysis. *Found. Trends Inf. Retr.* **2008**, *2*, 1–135. [CrossRef]

21. Brob, J. Aspect-Oriented Sentiment Analysis of Customer Reviews Using Distant Supervision Techniques. Ph.D. Thesis, University of Berlin, Berlin, Germany, 2013.

22. Bjorkelund, E.; Burnett, T.H.; Norvag, K. A study of opinion mining and visualization of hotel reviews. In Proceedings of the 14th International Conference on Information Integration and Web-based Applications & Services, Bali, Indonesia, 3–5 December 2012; pp. 229–238. [CrossRef]

23. Gräbner, D.; Zanker Fliedl, M.G.; Fuchs, M. Classification of customer reviews based on sentiment analysis. In *Information and Communication Technologies in Tourism*; Springer: New York, NY, USA, 2012; pp. 460–470. [CrossRef]

24. Hutto, C.; Gilbert, E. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In Proceedings of the 8th International AAAI Conference on Weblogs and Social Media, Ann Arbor, MI, USA, 1–4 June 2014.

25. Misopoulos, F.; Mitic, M.; Kapoulas, A.; Karapiperis, C. Uncovering customer service experiences with Twitter: The case of airline industry. *Manag. Decis.* **2014**, *52*, 705–723. [CrossRef]

26. Claster, W.B.; Pardo, P.; Cooper, M.; Tajeddini, K. Tourism, travel and tweets: Algorithmic text analysis methodologies in tourism. *Middle East J. Manag.* **2013**, *1*, 81–99. [CrossRef]

27. Kasper, W.; Vela, M. Sentiment analysis for hotel reviews. In Proceedings of the Computational Linguistics-Applications Conference, Jachranka, Poland, 17–19 October 2011; Volume 231527, pp. 45–52.

28. Jeon, W.; Lee, Y.; Geum, Y. Airline Service Quality Evaluation Based on Customer Review Using Machine Learning Approach and Sentiment Analysis. *J. Soc. e-Bus. Stud.* **2021**, *26*, 15–36.

29. Duan, W.; Cao, Q.; Yu, Y.; Levy, S. Mining online user-generated content: Using sentiment analysis technique to study hotel service quality. In Proceedings of the 46th System Sciences (HICSS), Hawaii International Conference, Maui, HI, USA, 7–10 January 2013; pp. 3119–3128. Available online: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6480220 (accessed on 12 April 2023).

30. Park, E.; Kang, J.; Choi, D.; Han, J. Understanding customers' hotel revisiting behaviour: A sentiment analysis of online feedback reviews. *Curr. Issues Tour.* **2020**, *23*, 605–611. [CrossRef]

31. Yu, Y.; Wang, X. World Cup 2014 in the Twitter World: A big data analysis of sentiments in US sports fans' tweets. *Comput. Hum. Behav.* **2015**, *48*, 392–400. [CrossRef]

32. Strapparava, C.; Mihalcea, R. SemEval-2007 task 14: Affective text. In Proceedings of the 4th International Workshop on Semantic Evaluations (SemEval '07), Association for Computational Linguistics, Prague, Czech Republic, 23–24 June 2007; pp. 70–74.

33. Holzman, L.E.; Pottenger, W.M. Classification of Emotions in Internet Chat: An Application of Machine Learning Using Speech Phonemes. Available online: https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=420ddd737833e808965e051a8af20d8bcd76f423 (accessed on 12 April 2023).

34. Alm, C.O.; Roth, D.; Sproat, R. Emotions from Text: Machine Learning for Text-based Emotion Prediction. In Proceedings of the Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing, Vancouver, BC, Canada, 6–8 October 2005; pp. 579–586.

35. Aman, S.; Szpakowicz, S. Identifying Expressions of Emotion in Text. Available online: http://saimacs.github.io/pubs/2007-TSD-paper.pdf (accessed on 12 April 2023).

36. Ding, X.; Liu, B.; Yu, P.S. A holistic lexicon-based approach to opinion mining. In Proceedings of the International Conference on Web Search and Web Data Mining, Palo Alto, CA, USA, 11–12 February 2008; pp. 231–240. [CrossRef]

37. Buhalis, D. Marketing the competitive destination of the future. *Tour. Manag.* **2000**, *21*, 97–116. [CrossRef]

38. Chen, J.; Becken, S.; Stantic, B. Assessing destination satisfaction by social media: An innovative approach using Importance-Performance Analysis. *Ann. Tour. Res.* **2022**, *93*, 103371. [CrossRef]

39. Tourism Research Australia. Chinese Visitor Satisfaction. 2014. Available online: https://www.tourism.australia.com (accessed on 12 April 2023).

40. Maedche, A.; Staab, S. Ontology Learning for the Semantic Web. *IEEE Intell. Syst.* **2001**, *16*, 72–79. [CrossRef]

41. Ganu, G.; Elhadad, N.; Marian, A. Beyond the Stars: Improving Rating Predictions Using Review Text Content. In Proceedings of the 12th International Workshop on the Web and Databases (WebDB 2009), Providence, RI, USA, 28 June 2009. Available online: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.150.140&rep=rep1&type=pdf (accessed on 12 April 2023).

42. Wan, H.; Yang, Y.; Du, J.; Liu, Y.; Qi, K.; Pan, J.Z. Target-Aspect-Sentiment Joint Detection for Aspect-Based Sentiment Analysis. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; Volume 34, pp. 9122–9129. [CrossRef]

43. Pontiki, M.; Galanis, D.; Papageorgiou, H.; Manandhar, S.; Androutsopoulos, I. Semeval-2015 task 12: Aspect based sentiment analysis. In Proceedings of the 9th International Workshop on Semantic Evaluation, Denver, CO, USA, 4–5 June 2015; pp. 486–495.

44. Pontiki, M.; Galanis, D.; Papageorgiou, H.; Androutsopoulos, I.; Manandhar, S.; Al-Smadi, M.; Al-Ayyoub, M.; Zhao, Y.; Qin, B.; Clercq, O.D.; et al. Semeval-2016 task 5: Aspect based sentiment analysis. In *ProWorkshop on Semantic Evaluation (SemEval-2016)*; Association for Computational Linguistics: Stroudsburg, PA, USA, 2016; pp. 19–30.

45. Li, S.; Scott, N.; Walters, G. Current and potential methods for measuring emotion in tourism experiences: A review. *Curr. Issues Tour.* **2015**, *18*, 805–827. [CrossRef]

46. D'Mello, S.; Picard, R.W.; Graesser, A. Toward an Affect-Sensitive AutoTutor. *IEEE Intell. Syst.* **2007**, *22*, 53–61. [CrossRef]

47. McLaren, K. (n.d.) Your Emotional Vocabulary List. Dynamic Emotional Integration. Available online: https://karlamclaren.com/wp-content/uploads/2016/05/Emotional-Vocabulary-List-Color.pdf (accessed on 12 April 2023).

48. Plutchik, R. *Emotion: A Psychoevolutionary Synthesis*; Harper and Row: New York, NY, USA, 1980.

49. Elbagir, S.; Yang, J. Sentiment Analysis on Twitter with Python's Natural Language Toolkit and VADER Sentiment Analyzer. In *IAENG Transactions on Engineering Sciences*; 2020; pp. 63–80.

50. Fleiss, J.L. *Statistical Methods for Rates and Proportions*, 2nd ed.; John Wiley: New York, NY, USA, 1981.

51. Altman, D.G. *Practical Statistics for Medical Research*; Chapman & Hall/CRC Press: New York, NY, USA, 1999.

52. Volo, S. Bloggers' reported tourist experiences: Their utility as a tourism data source and their effect on prospective tourists. *J. Vacat. Mark.* **2010**, *16*, 297–311. [CrossRef]

53. Dodds, P.S.; Clark, E.M.; Desu, S. Human language reveals a universal positivity bias. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 2389–2394. [CrossRef]

54. Vishnubhotla, K.; Mohammad, S.M. Tweet Emotion Dynamics: Emotion Word Usage in Tweets from US and Canada. In Proceedings of the 13th Language Resources and Evaluation Conference (LREC-2022), Marseille, France, 20–25 June 2022.

55. Pappas, N.; Popescu-Belis, A. Sentiment Analysis of User Comments for OneClass Collaborative Filtering over TED Talks. In Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval, Dublin, Ireland, 28 July–1 August 2013; ACM: New York, NY, USA; pp. 773–776.