Tourism Mobile App With Aspect-Based Sentiment Classification Framework for Tourist Reviews

Muhammad Afzaal, Muhammad Usman, and Alvis Fong, *Senior Member, IEEE*

***Abstract*—Tourist reviews are information sources for travelers to know about tourist places. Unfortunately, some reviews are irrelevant and become noisy data. Aspect-based sentiment classification methods have shown promise in suppressing the noise. However, not much research has been done on automatic aspect identification, and identification of implicit, infrequent and co-referential aspects, resulting in misclassifications. This paper presents a framework of aspect-based sentiment classification that will not only identify the aspects very efficiently but can perform classification task with high accuracy. The framework has been implemented as a mobile app that helps tourists find the best restaurant or hotel in a city, and performance has been evaluated by conducting experiments on real-world datasets with excellent results (85% identification and 90% classification).**

***Index Terms*—Machine learning, consumer reviews, consumer mobile app, aspect based sentiment analysis.**

I. Introduction

TOURISM is a growing industry with increasing importance to countries globally [1]. With the advent of smartphones, tourists visit places of interest and share their sentiments on various social platforms and websites. These sentiments present a broad view to readers about a tourist place. However, readers may be confused about whether to visit the tourist place due to diversity of sentiments. Sentiment classification methods [2]-[6] can help to organize the sentiments into positive and negative. However, each tourist place has diverse aspects and a simple binary result is often inadequate. So, aspect-based sentiment classification methods [7]-[20] can be useful. For example, an opinion "food is yummy, but place is dirty" has two aspects: "food" and "place". Food is classified as positive by the positive word "yummy"; place is classified as negative by the word "dirty". These methods involve two basic tasks: (1) aspects identification and (2) sentiment classification into positive or negative about identified aspects. Both are difficult tasks [21].

There are three problems associated with the first task of aspect identification: (1) identify the implicit aspects, e.g., "Yesterday we went to XX restaurant, every dish was oily, soggy and super salty." This statement implicitly gives an opinion about an important aspect "food" that was not mentioned in the original text.

(2) Identify the co-referential aspects, which refer to the same aspect that mentions in opinions with different word or synonyms, e.g., atmosphere and ambiance are co-referential aspects because both refer to the same aspect. (3) Identify the aspects that occur infrequently but are nonetheless important, e.g., curtains and bed are somewhat unusual aspects but important for hotels. In the second task of aspect-based sentiment classification about identified aspects, the main problem is irrelevant sentences like self-introduction, details about the previous visit, etc. that affect the accurate prediction about aspects of a tourist place.

This paper presents an effective framework of aspect-based sentiment classification by introducing state-of-the-art machine learning algorithms. The framework consists of two main elements (1) decision tree-based aspect identification method, which allows readers to identify explicit, implicit and infrequent aspects, and groups co-referential aspects from tourist sentiments (2) aspect-based sentiment classification using machine learning algorithms that has three stages. In the first stage, Stanford Basic Dependency method [22] is used to filter sentence parts between sentiment words and aspects in a given sentiment sentence. In the second stage, filtered phrases are used to build features like n-grams and Part-Of-Speech tags. Lastly, machine learning algorithms are applied to identify features to classify the opinions about aspects into positive or negative. Evaluation is performed using 10-fold cross-validations to limit problems like overfitting [23].

This paper aims to identify and classify the aspects mentioned in tourist sentiment using machine learning algorithms. These are very common in daily life where decisions are taken by interlinked multiple criteria. In the proposed framework, machine learning classifies a given sentiment by various aspects and sentiment words, e.g., in a restaurant review, reviewer likes the food but dislikes the service. The algorithm classifies the review by sentiment words or phrases about aspects.

Real-world hotel and restaurant reviews were collected from travel websites for evaluating the effects of dataset size, time, feature size, feature types and feature weighted methods on the performance of the proposed framework. The experiments show that aspect-based sentiment classification framework is very effective. Five machine learning algorithms are evaluated and compared. They are Naive Bayes Multinomial (NBM) [24], Support Vector Machine (SVM) [25], Maximum Entropy (ME) [26], Random Forest Tree (RFT) [27], and Fuzzy Lattice Reasoning (FLR) [28]. These machine

Manuscript received August 11, 2018; revised December 3, 2018 and February 27, 2019; accepted March 29, 2019. Date of publication April 2, 2019; date of current version April 23, 2019. *(Corresponding author: Alvis Fong.)*

M. Afzaal and M. Usman are with the Department of Computer Science, Shaheed Zulfikar Ali Bhutto Institute of Science and Technology, Islamabad 44000, Pakistan (e-mail: ziafigover@gmail.com; dr.usman@szabist-isb.edu.pk).

A. Fong is with the Department of Computer Science, Western Michigan University, Kalamazoo, MI 49008 USA (e-mail: acmfong@gmail.com).

Digital Object Identifier 10.1109/TCE.2019.2908944

learning algorithms are also compared with other related methods [7], [11], [18], [20]. Comparison results show considerably improved performance by using the proposed framework with 88.08% and 90.53% accuracy on restaurant dataset and hotels dataset, respectively, with NBM. In addition, the proposed aspect-based sentiment mining framework has been implemented as a mobile app. It allows users to view tourist sentiments by aspects.

The rest of the paper is organized as follows. Section II presents an overview of aspect-based sentiment mining. Section III discusses aspect identification and classification models. Section IV presents experimental results on real-world datasets and comparative evaluation of the proposed framework with related methods. Sections V and VI present the mobile app and its analysis. Finally, Section VII concludes the paper.

II. Related Work

This section reviews related work of aspect-based sentiment classification in the tourism domain. The two areas of focus are: (1) Aspect Identification, and (2) Aspect-based Sentiment Classification.

A. Aspect Extraction

In aspect-based sentiment classification, aspect identification is the most important task. Existing methods come under three main categories: rule based, seeds based, and topic models based [21]. Rule-based methods often rely on extraction rules derived from importance, frequency, and appearance. Marrese-Taylor *et al.* [8]-[9] propose a two-phase method that first applies Part-Of-Speech Tagger on sentences, followed by extraction of nouns as aspects. Hai *et al.* [10] use intrinsic domain relevance (IDR) and extrinsic domain relevance (EDR) score for tourism domain. Afzaal *et al.* [11] propose an improved rule-based method using FURIA machine learning algorithm. However, FURIA does not build enough rules to extract sufficient aspects, and infrequent aspects are not extracted.

Seeds based methods identify aspects using the grammatical connection between seeds words with sentiment words [12]-[13]. Bootstrapping has been applied to quantify overlapping and dependency between seed word and review words [14]. Zhu *et al.* [15] use bootstrapping to identify important aspects. They consider two kinds of terms for aspects distinguishing proof (1) Part-Of-Speech Tagger and (2) N-Grams. Like rule-based methods, seeds-based techniques lack the ability to identify infrequent aspects.

Topic model-based techniques assume that every sentiment is a blend of different topics, and every point is a probability distribution over various words, e.g., Wu and Ester [16] assume that every sentiment about restaurant and hotel is associated with an aspect, and they use added substance generative techniques to identify the aspects. Xianghua *et al.* [17] apply a sliding window to identify the aspects from tourist reviews, but the initial step of some of the aspects is inaccurate. Xu *et al.* [18] apply Latent Dirichlet Allocation to topic modeling, but the confinements of theme model-based strategies administered by "co-event" (i.e. how regularly terms co-occur in various settings) lead to numerous "non-particular" and "immaterial" viewpoints being pulled and bunched.

The main limitations of the above methods are not removing irrelevant aspects and not managing co-referential aspects issue in reviews. The latter means individuals utilize different expressions to describe a similar aspect. Moreover, there are aspects that do not show up straightforwardly in reviews but rather the sign of the review to a specific aspect. For instance, "food" is an implicit aspect in "The taste is great, too".

B. Aspect-Based Sentiment Classification

Aspect-based sentiment classification attempts to determine of the orientation of views of the given text in two or more classes (good, bad or five stars) about aspects. Wang *et al.* [14] use Latent Rating Regression (LRR) to organize relevant words into one of the five different ratings on the rating scale. Xu *et al.* [18] predict opinion about specific aspects like "Staff", "Food" and "Ambiance" using SVM on real world datasets after preprocessing with NLP toolkit for sentence segmentation and POS Tagger. Pontiki *at el*. [19]-[20] use SVM with linear kernel (SV classifier) to perform binary classification with results that indicate robustness and stability. De Albornoz *et al.* [7] represent survey rating as Vector of Feature Intensities (VFI) and used lodging audits to validate their method. Afzaal *et al.* [11] propose a three-stage fuzzy aspect-based classification method to classify sentiments into positive and negative.

A limitation of supervised machine learning is that the reviews data must be labeled to train the classifier. Some data already come with labels assigned by reviewers. Reviewer-labeled data can be noisy, with unwanted details like self-introduction, previous stories, etc. that negatively affect the classification. Otherwise, manual labeling, which is tedious and expensive, is required.

III. Proposed Framework

Fig. 1 presents an overview of the proposed framework for aspect identification and classification. In Step 1 (Data Collection), tourist reviews about tourist places like hotels and restaurants are collected from multiple social media platforms and websites. Step 2 (Data Preprocessing), suppresses noise and redundancy, and cleaned reviews are transformed into sentences. Step 3 (Aspect Identification) finds aspects from preprocessed datasets using a hybrid aspect identification method. Step 4 (Classification) uses machine learning to classify the identified aspects into positive or negative sentiment.

A. Data Collection

In data collection, reviews are collected from popular social media websites using crawler and APIs. The datasets have different numbers of reviews in each domain. In the restaurant domain, there are 2000 reviews with 1000 positive and 1000 negative. In the hotel domain, there are 4000 reviews with 2000 positive and 2000 negative. London is chosen as a city of interest in the case study.

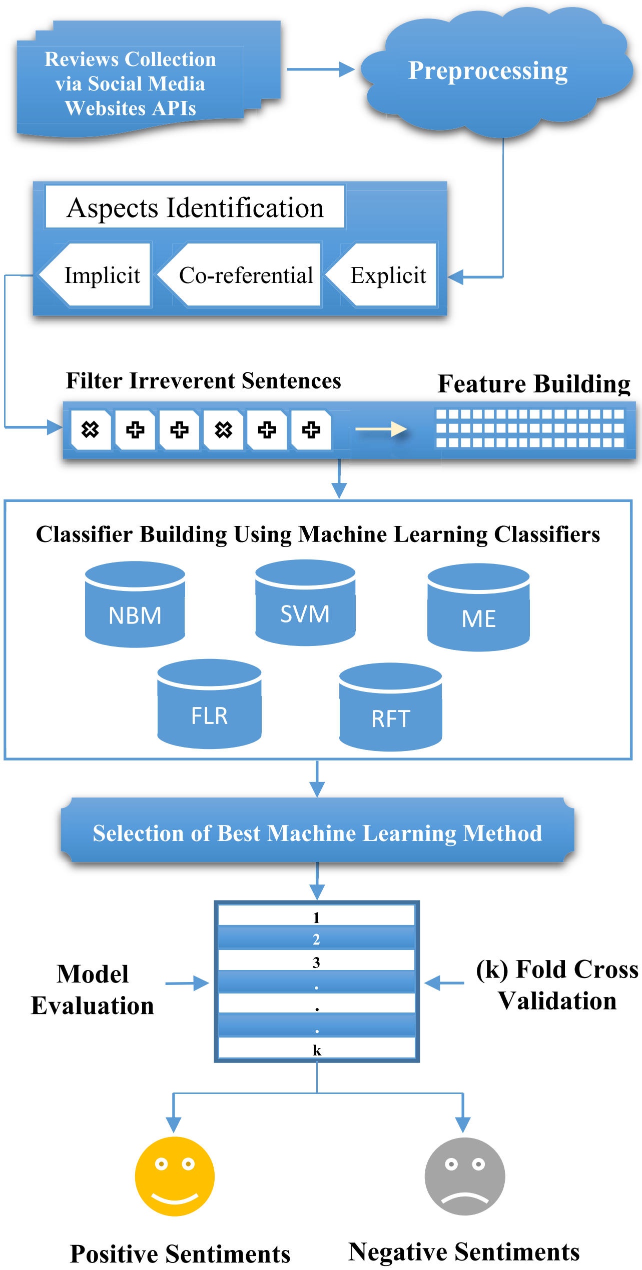


Fig. 1. Framework for aspect-based sentiment classification.

Algorithm 1 Hybrid Tree Based Aspect Identification

Input: Collection of sentences = $\{S_1, S_2, S_3 \ldots S_n\}$
Output: Aspects assigned to sentences

1. **initialize** aspects
2. **for** all sentences **do**
3. stanford tagger = SPOS (sentences)
4. **if** NN in stanford tagger **then**
5. aspects ← NN
6. **end if**
7. **end for**
8. **initialize** aspects groups
9. **for** all aspects **do**
10. WordNet_sets = WNSS (aspects)
11. **if** TRUE in WordNet sets **then**
12. aspects_groups ← aspects
13. **end if**
14. **end for**
15. frequent_aspects = freq_measure (aspects, aspects_groups, 10)
16. tree = DT (sentences, frequent aspects)
17. **initialize** aspect_assigned_sentences
18. **for all** sentences **do**
19. aspect_identification = tree (sentences)
20. **if** TRUE in aspect_identification **then**
21. aspect_assigned_sentences ← aspect identification
22. **end if**
23. **end for**
24. **return** aspect_assigned_sentences

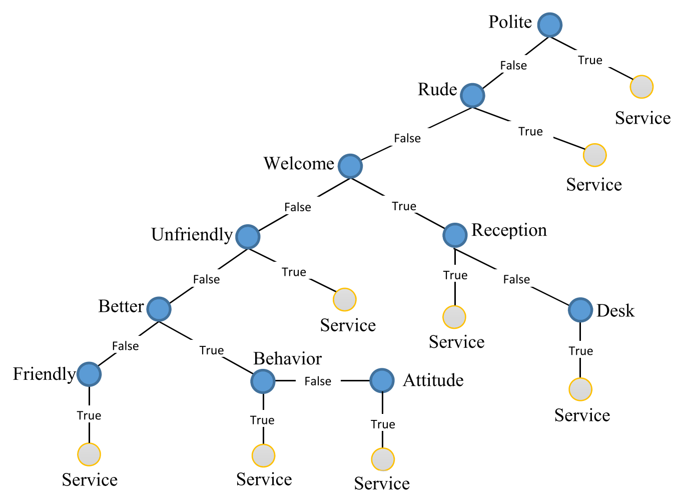


Fig. 2. Decision tree for service aspect identification in restaurants data.

B. Data Preprocessing

Data preprocessing removes redundancy and ambiguity inherit in the data and transforms the reviews into sentences to facilitate sentence-level aspect-based classification. First, sentences are extracted by identifying the delimiters (e.g., dot, exclamation or question mark). Next, redundant information, e.g., duplicate sentences, is removed. Finally, ambiguous, vague or misspelled terms are corrected. The cleaned restaurant and hotel datasets contain 3787 and 7802 sentences, respectively.

C. Aspects Identification

The objective of aspect identification method is to identify aspects that are important and relevant to a tourist place. This paper proposes a hybrid aspect identification method that can identify both explicit and implicit aspects from reviews about tourist places. Algorithm 1 below describes the workflow of this approach. The algorithm takes all sentiment sentences as input for aspect identification, and then processes the input sentences and assigns relevant aspect to each sentence as output.

Lines 1-7 of Algorithm 1 identify explicit aspects from sentiment sentences. Stanford Part-Of-Speech Tagger [29] is applied to each sentence to obtain a lexicon of POS tags (Lines 2 and 3). Then, all POS tags are discarded except the Noun (NN) and Noun Phrases to be used as explicit aspects (Lines 4-6). Lines 8-14 group all co-referential aspects with the same meaning or point to the same aspect (e.g., location, place, and venue), and synonyms (e.g., atmosphere and

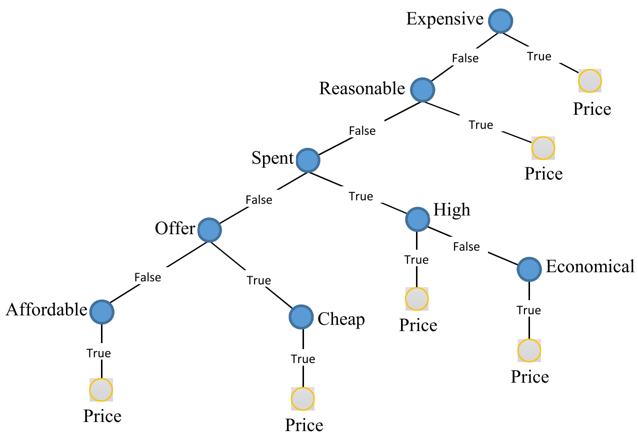


Fig. 3. Decision tree for price aspect identification in hotels data.

environment) using WordNet [30]. Line 15 applies a rule base to extract relevant aspects with >10 occurrences in all sentences. At this stage, most sentences have assigned explicit aspects. Next, for implicit aspects identification, decision trees are used for the restaurant and hotel domains as shown in Figs. 2 and 3. Sentences are segmented into words using Uni-Gram Tokenizer as input to the respective decision trees. Line 16 shows the use of words as condition and assigned aspect as a class. Lines 17-23 then find the implicit aspects from the decision trees. Finally, Line 24 returns all the assigned aspects.

D. Aspect-Based Opinion Classification

The classification process consists of three basic stages. First, sentences that have no sentiment about assigned aspect are discarded. Second, features are extracted from sentiment sentences using N-Grams and POS. Third, classifiers are trained on the extracted features.

Discard Opinion Sentences: Many sentences in reviews are irrelevant as they do not contain sentiment information, e.g., "I ordered fish, chicken, and dessert." Stanford Basic Dependency [22] algorithm was applied to each review sentence to find dependencies between sentiment words (adjective) and aspect. If an aspect does not have any dependency with sentiment words, then it will be removed from the sentences.

Feature Extraction: Four types of features are extracted from each dataset. They are N-Grams (N = 1, 2, 3) and POS tags, which are obtained in three steps. The first step is tokenization, which for POS tags only verb, adverb, and adjective are extracted from the dataset. Second, stop words are removed. Finally, negation is correctly processed, e.g., "I do not like rice" will form three bigrams: "I do+not," "do+not like," "not+like rice."

Classifier Training: A machine learning algorithm classifies each aspect in a consumer review into positive or negative by considering all aspects and their linkages to sentiment words. For example, in a restaurant review, the tourist likes the food but dislikes the service. The class of this review depends on the sentiment words and phrases linked to aspects. When multiple aspects are considered, the situation becomes more complex; machine learning algorithms are very efficient and helpful. The five popular machine learning algorithms that are considered in this research are described below.

Naïve Byes Multinomial: Naïve Bayes classifiers are simple probabilistic classifiers derived from an application of Bayes' theorem but with strong independence assumptions between the features. The independence assumptions mean that the probability of observing features from f_1 to f_n, for a given class c, can be computed as a simple product of the form [24]:

$$p(f_1,\ldots,f_n|c) = \prod_{i=1}^{n} p(f_i|c) \tag{1}$$

This means that when a new example is classified using Naive Bayes, it becomes simpler to work with posterior probability of the form [24]:

$$p(c|f_1,\ldots,f_n) \propto p(c)p(f_1|c)\cdots p(f_n|c) \tag{2}$$

Although the independence assumptions rarely, if ever, hold in practice, Naive Bayes classifiers have been found to perform well even for some complicated tasks. In addition, Naive Bayes classifiers are scalable as they require a number of parameters linear in the number of features. A logical extension of the Naïve Bayes model is Multinomial Naive Bayes, which allows each feature distribution $p(f_i|c)$ to be a multinomial distribution. This method works well for easily countable data like words in text.

Support Vector Machine: SVM attempts to separate data points in a multidimensional feature space by defining decision boundaries with hyperplanes. Support vectors are data points closest to a decision boundary in the feature space. Although SVM can be used in multiclass classification problems, it is fundamentally a binary classification method in which an unseen document vector is assigned to one of two classes separated by a hyperplane. Given a training set with labeled pair $(x_i, y_i), i = 1, 2, \ldots$ where $x_i \in R^n$ and $y \in \{1, -1\}^l$, SVM entails an optimization problem of the form [25]:

$$\begin{aligned} &\min_{w,b,\mathcal{E}} \ \frac{1}{2}W^T W + C\sum_{i}^{l}\mathcal{E}_i \\ &\textit{Subject to}\ \ y_i\left(w^T \emptyset(X_i) + b\right) \geq 1 - \mathcal{E}_{i,} \\ &\qquad\qquad\ \ \mathcal{E}_i \geq 0 \end{aligned} \tag{3}$$

where W contains weights assigned to features, $\mathcal{E}$ models added error correction and C is a regularization parameter. The optimization problem is one of minimizing $\frac{1}{2}W^T W + CP\sum_{i=1}^{l}\mathcal{E}_i$, where value of $y_i(w^T\emptyset(X_i)+b)$ needs to be greater than $1-\mathcal{E}_i$ and $\mathcal{E}$ needs to approach zero. Training vector x_i is mapped to higher dimensional space by $\emptyset$. In the present context, the written reviews must be converted to numeric vectors. The vectors are then normalized to keep them in range between 0 and 1.

Maximum Entropy: This algorithm uses conditional distribution constraints to model the training data's characteristics. The Maximum Entropy (ME) value is given by [26]:

$$P_{ME}(c|d) = \frac{1}{Z(d)}\exp\left(\sum_i \lambda_{i,c} f_{i,c}(d,c)\right) \tag{4}$$

where $P_{ME}(c|d)$ is the probability of document d being classified as class 'c', $f_{i,c}(d, c))$ is the feature/class function for f_i

feature and c class, $\lambda_{i,c}$ is the estimating parameter, and Z(d) is a normalization factor. To classify unseen text, count of the words is taken as a feature. The feature/class function can be instantiated as [26]:

$$f_{i,c}(d,c) = \begin{cases} 0 & if \quad c \neq c' \\ \frac{N(d,i)}{N(d)} & otherwise \end{cases} \tag{5}$$

where $f_{i,c}(d,c)$ contains the features in word-class combination of class c and document d, N(d,i) represents occurrence of feature i in document d, and N(d) is the word count in d. Thus, each word has its word-class pair and its associated weight. If a word occurs frequently in a class, then that word gets a higher weight of relevant word-class pair as compared to others. The word-class pairs with the highest frequency are used in classification.

Random Forest [27], [31]: Individual decision trees do not generally give good classification performance because of high variance, even though bias can be low. Bagging is often applied to build multiple decision trees to reduce variance without significantly affecting bias. Bagging trains each tree on a bootstrap sample of training data; a majority vote of decision trees is then used for making predictions to even out the variance. Different from bagging, random forests do not use the full set of features as candidates for splitting at each decision node. Instead, only a randomly selected subset of features (typically square root of the total number of features) is split at each node of growing tree. This has an effect of further reducing variance.

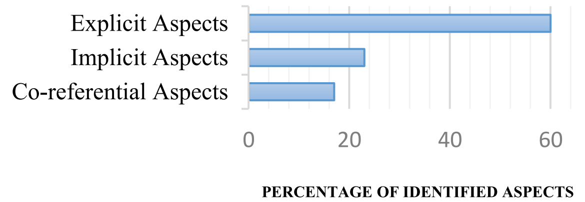
Fuzzy Lattice Reasoning: This method attempts to find fuzzy lattice rules from the training data, and then use these rules to classify unseen data [28]. Suppose U is the universal set of data objects of all types. A fuzzy lattice is written as $<L, u>$, where L is a lattice and u is its valuation function. The lattice comprises many elements, and each element belongs to a certain class c_i, where each c_i belongs to the set C. Each fuzzy lattice rule is a couple $<ui, ci>$, where u_i is an object and c_i is its related class that performs the function $h : U{-}{>}C$. These couples are derived from the whole training set. When a new object comes in the test set for which no rule is available in training set, then the existing rules compete to classify it. The new object is assigned to one of the classes of set C based on the inclusion measure parameter of the object.

IV. Evaluation

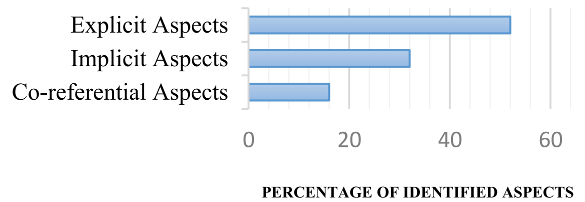
This section presents experimental results of applying the aspect-based sentiment classification framework to reviews taken from popular tourist websites.

A. Aspects Identification

Both the restaurants and hotels datasets were used to measure the proportion of correctly identified aspects in each dataset. The proposed aspects identification method correctly identified 80% of the aspects in the restaurants dataset and 85% of the aspects in the hotels dataset. A breakdown of the types of aspects found by the identifier is also of interest. Fig. 4 shows a breakdown of the types of aspects found in the restaurants and hotels datasets. In the restaurants dataset, the most common aspects identified were explicit (frequent and infrequent). They made up 60% of all identified aspects. The second were implicit aspects with 23%, while co-referential aspects made up 17% of all identified aspects. Similarly, in the hotels dataset, explicit aspects (frequent and infrequent) were most commonly found with 52%. Next were implicit aspects with 32%, and co-referential aspects were again third with 16%.



(a)



(b)

Fig. 4. Correctly identified aspects. (a) On restaurants dataset. (b) On hotels dataset.

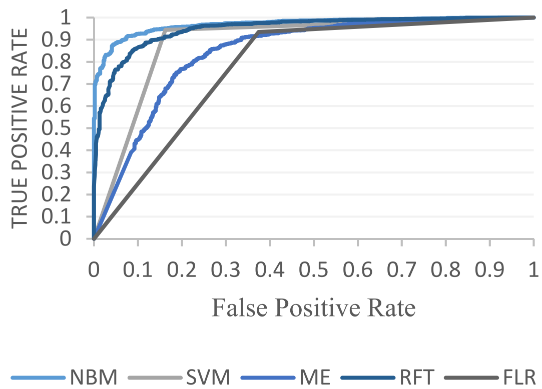
TABLE I
CLASSIFIERS PERFORMANCE

Classifier	Accuracy	Precision	Recall	F-Measure
		Restaurants Dataset		
NBM	**88.08%**	**0.88**	**0.88**	**0.88**
SVM	87.02%	0.87	0.87	0.87
ME	85.74%	0.85	0.85	0.85
RFT	87.42%	0.87	0.87	0.87
FLR	77.96%	0.78	0.78	0.78
		Hotels Dataset		
NBM	**90.53%**	**0.90**	**0.90**	**0.90**
SVM	89.93%	0.89	0.89	0.89
ME	83.83%	0.84	0.84	0.84
RFT	87.79%	0.88	0.88	0.87
FLR	80.12%	0.80	0.80	0.80

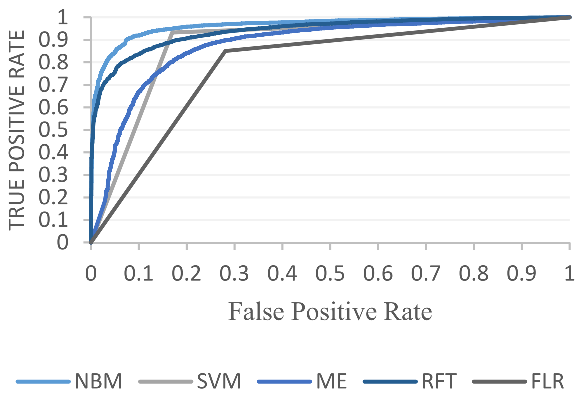
B. Aspect-Based Sentiment Classification

To measure the aspect-based sentiment classification performance, each algorithm was evaluated on varied sizes of the dataset, different feature weighting methods, and different feature types. The classification time on varied sizes of datasets was also measured to gauge the relative latency of each method. The results are shown in Table I and Figs. 5-9. Table I shows aspect-based sentiment classification accuracy of the five machine learning methods under evaluation. For both datasets, the best performance was achieved using NBM; it achieved 88.08% classification accuracy on the restaurants dataset and 90.53% on the hotels dataset.

Figure 5 shows the ROC curves plotted based on the result of aspect-based sentiment classification. It again shows that NBM performed the best among the five machine learning algorithms on both datasets. Fig. 6 shows the time taken

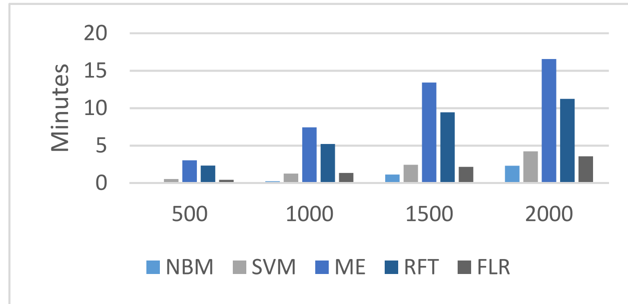


(a)

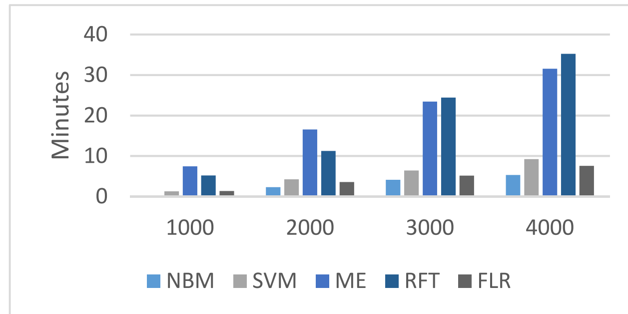


(b)

Fig. 5. ROC curves. (a) On restaurants dataset. (b) On hotels dataset.

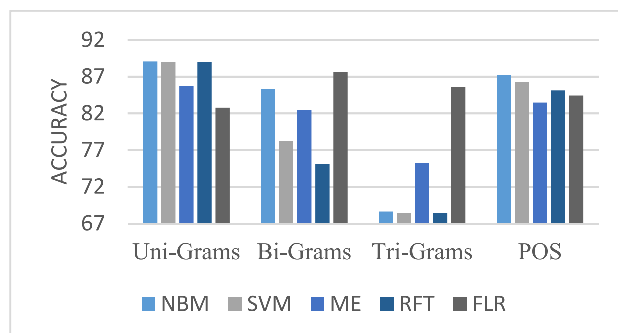


(a)

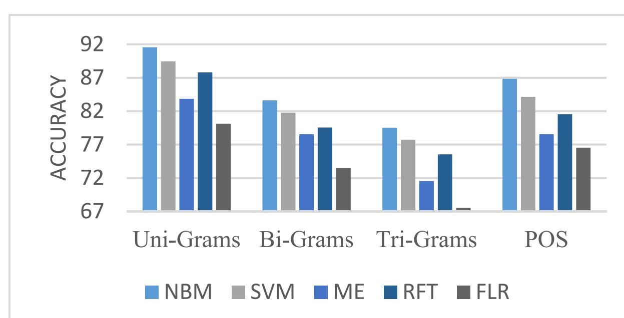


(b)

Fig. 6. Classifiers prediction time with different instance sizes. (a) On restaurants dataset. (b) On hotels dataset.



(a)



(b)

Fig. 7. Classifiers performance with different feature types. (a) On restaurants dataset. (b) On hotels dataset.

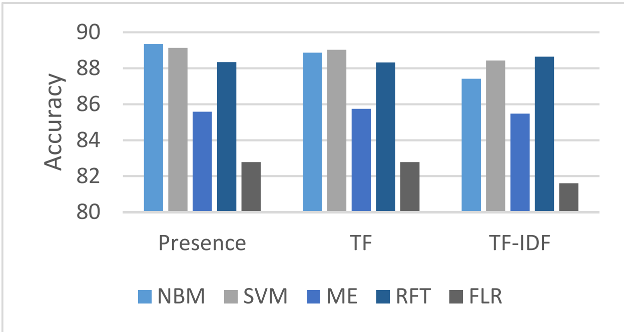
to predict labels of the reviews. NBM was the fastest on both datasets and FLR was the slowest, highlighting the complexity of the method associated with building and applying fuzzy lattice rules. Fig. 7 shows how each of the four feature types (Unigrams, Bigrams, Trigrams, and POS) affected the aspect-based sentiment classification performance. It shows that Unigrams and POS worked best with NBM on both datasets. Fig. 8 shows how feature weighting methods, such as Presence, TF, and TF-IDF, affected the aspect-based sentiment classification performance. It shows that Presence worked best with NMB on both datasets.

Figure 9 shows how different dataset sizes affected the aspect-based sentiment classification performance. The restaurant reviews were split into four chunks of 500, 1000, 1500 and 2000, and the hotel reviews were split into four chunks of 1000, 2000, 3000 and 4000. Using NBM, the best results were achieved using 1000 reviews chunk for the restaurants dataset, whereas for the hotels dataset 4000 reviews chunk was best.

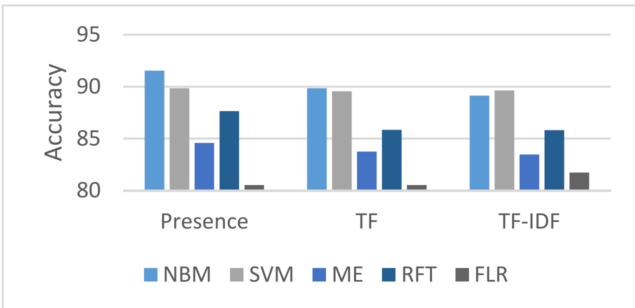
C. Comparative Evaluation

The novel machine learning framework was then compared with existing methods in the tourism domain. Table II and Table III present comparisons with respect to aspects identification and aspect-based sentiment classification, respectively. The results show that the proposed framework outperforms others in both tasks.

The proposed method excels in computational complexity and performance. Logistic classification has low complexity but performed poorly as shown in Table III. SVM performed well but has high complexity. Similarly, although Afzaal *et al.*

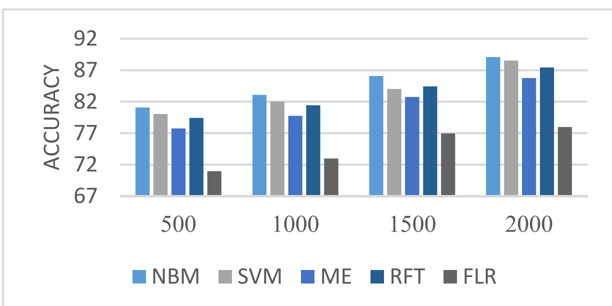


(a)

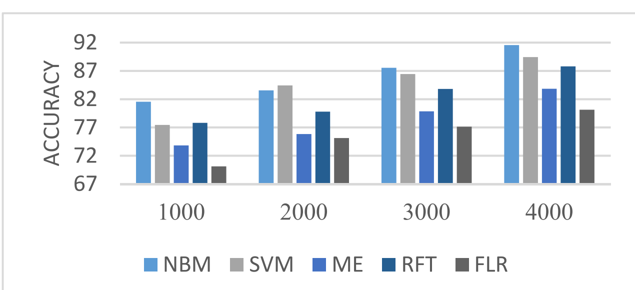


(b)

Fig. 8. Classifiers performance with different feature methods. (a) On restaurants dataset. (b) On hotels dataset.



(a)



(b)

Fig. 9. Classifiers performance with different instance sizes. (a) On restaurants dataset. (b) On hotels dataset.

achieved accurate results [11], it relied on highly complex 3-stage fuzzy classifiers. In contrast, Naïve Bayes has low complexity (all it needs to do is to compute the frequency of each feature value for each class), and the multinomial extension does not add much in complexity. At the same time, NBM achieved superior results compared with others.

TABLE II
COMPARISON: ASPECTS IDENTIFICATION

Reference	Explicit Aspects		Implicit Aspects	Co-referential Aspects?	Discard Irrelevant Aspects?	Method	Result %
	Frequent	Infrequent					
de Albornoz, J. C. et al., 2011 [7]	High	Null	Null	Y	Y	Rules Based	66.8
Mukherjee, A. et al., 2012 [13]	High	Low	Null	N	Y	Seeds Based	77
Xianghua, F. et al., 2013 [17]	High	Medium	Null	N	N	LDA Based	73
Afzaal, M. et al., 2016 [11]	High	Low	Low	Y	N	FURIA Based	79
Proposed (Restaurants Dataset)	**High**	**Medium**	**High**	**Y**	**Y**	**Hybrid Tree Based**	**80**
Proposed (Hotels Dataset)	**High**	**Medium**	**High**	**Y**	**Y**	**Hybrid Tree Based**	**85**

TABLE III
COMPARISON: ASPECT-BASED SENTIMENT CLASSIFICATION

Reference	Data size	Method	Result %
de Albornoz, J. C. et al., 2011 [7]	1500	Logistic	71.7
Xueke, X. et al., 2013 [18]	3214	Support Vector Machine	83.9
Pontiki, M. el al., 2015 [20]	320	Maximum Entropy	78.69
Afzaal, M. et al., 2016 [11]	4000	Fuzzy Logical Reasoning	86.02
Proposed (Restaurants Dataset)	**2000**	**Naïve Bayes Multinomial**	**88.08**
Proposed (Hotels Dataset)	**4000**	**Naïve Bayes Multinomial**	**90.53**

V. MOBILE APP IMPLEMENTATION

The proposed framework has been implemented as a mobile app using open source technologies. The architecture of the mobile app is presented in Figure 10. It is divided into two sides, namely mobile side and server side, and services are being used as a communication link to synchronize data among both sides. Synchronization allows the transfer of data according to user preferences to avoid storing large amounts of unnecessary data on the mobile device. Mobile middleware is the main processing unit where algorithms are implemented for local aspect-based sentiment analysis by utilizing user's rich information that mobile phone already has. It is also responsible for synchronization of data between local and server storage. The advantage of this setup is that it does not only optimize the consumer's experience by providing useful information in a timely manner but also can operate on this information even when the mobile device is out of range or disconnected.

The mobile app encompasses the situation in London, where tourists use other tourist tools to find the best restaurants and hotels in the city. The working of the mobile application is presented in Fig. 11. Fig. 11(a) shows the location icons of restaurants and hotels on London city map with distinct colors. By clicking location icon of any restaurant or hotel, the app is

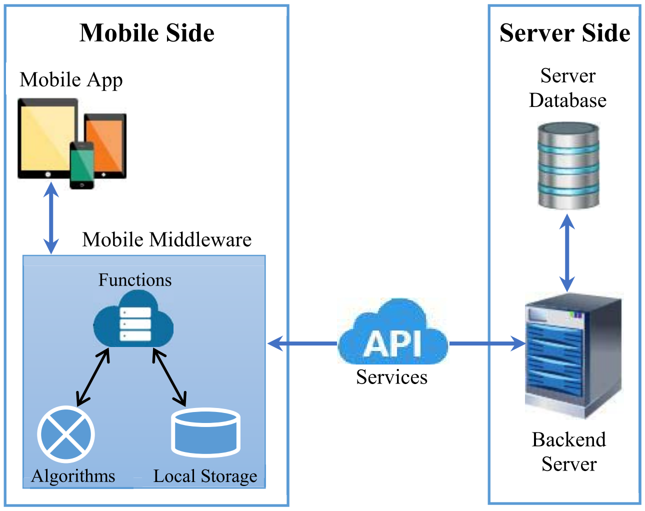


Fig. 10. Mobile app architecture.

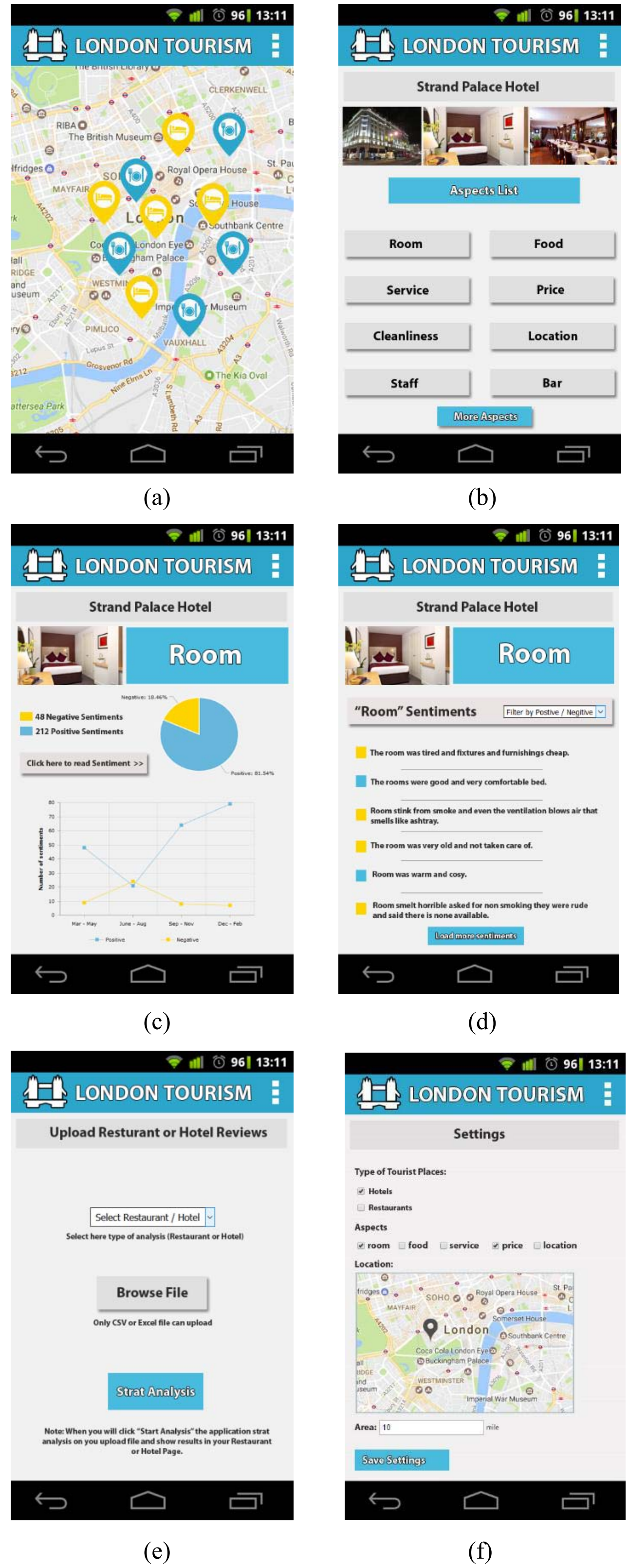


Fig. 11. Mobile app user interfaces. (a) Selected locations. (b) Hotel details. (c) Aspect details. (d) Sentiments page. (e) Analysis page. (f) Settings page.

redirected to hotel or restaurant detail page. Fig. 11(b) shows the hotel details page with hotel name, images, and corresponding aspects list. The list shows the top eight aspects plus a load more aspects button. By selecting any aspect, the app is redirected to the aspect details page as shown in Fig. 11(c).

Figure 11(c) shows useful information about the selected aspect. A pie chart shows the percentages of positive and negative sentiments about the room aspect. At the bottom of this page, a line chart divides the sentiments into four periods (Mar-May, June-Aug, Sep-Nov, and Dec-Feb), and shows the number of positive and negative sentiments in each period about the room aspect. At the left side, the total page number of positive and negative sentiments is shown with a button "click here to read sentiments." By clicking on this button, the application is redirected to sentiments detail page of the room aspect as shown in Fig. 11(d). Tourists can read sentiments regarding room aspect on this page. They can also apply the filter if they only want negative or positive sentiments of this aspect.

Figure 11(e) shows the interface for tourist organizations if they want to analyze their own restaurants or hotels reviews. They can upload unlabeled dataset in spreadsheet format and make an analysis. By clicking on the "start analysis" button, the mobile app automatically identifies aspects and classifies sentiments into positive or negative using the proposed framework. After completing the analysis, mobile app quickly displays the results, which is near-instantaneous and therefore suitable for consumer applications.

Figure 11(f) shows the settings screen that allows the user to set the app according to their preferences. Two preferences are introduced in this app such as "Tourist places and aspects-based preference" and "location-based preference". In tourist places and aspects-based preference, users can select which type of tourist places they want to view on the map and which aspects they want to analyze. For example, if a user selects hotels as tourist places and wants to analyze only price and room aspects, the mobile app will display hotels on the map and analysis of each hotel regarding selected aspects of price and room. In location-based preference, users have the option to view only those tourist places that are near their current location and they have control to set the area size in terms of miles.

TABLE IV
THE DEVICES USED IN THE EXPERIMENTS AND THEIR CHARACTERISTICS

	Year	Hard Disk	RAM	CPU	Battery (mAh)
Device 1	2018	16 GB	2 GB	1.4 GHz Quad Core	3000
Device 2	2018	8 GB	1 GB	1.2 GHz Quad Core	1200
Device 3	2015	4 GB	512 MB	1.2 GHz Quad Core	2000

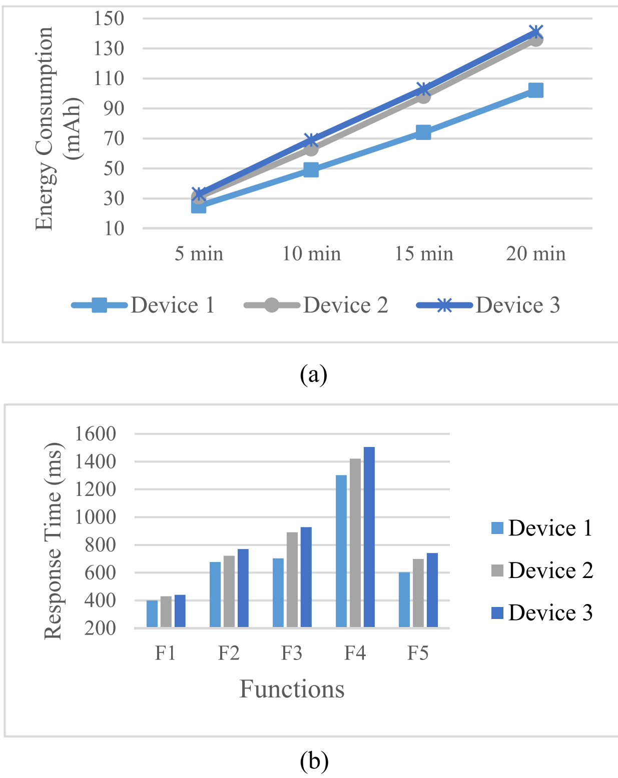


Fig. 12. Mobile app Analysis. (a) Device energy consumption using the app. (b) App performance on different devices.

VI. APPLICATION ENERGY CONSUMPTION AND PERFORMANCE ANALYSIS

This section presents an analysis of the mobile app in terms of energy consumption and performance. The ideas of mobile app analysis have come from different sources. Oliveira *et al.* [32] proposed a method for comparing energy consumption of different devices. Phan *et al.* [33] proposed a benchmark for mobile app performance analysis. To achieve diversity three devices were used in this analysis from different manufacturers, years of manufacture, and with different RAM, and CPU specifications as reported in Table IV. Manufactured in 2015, Device 3 represents a legacy device that demonstrate backward compatibility of the app. Fig. 12 (a) shows the result of energy consumption while the mobile app is in use on different devices.

The mobile app was run on each device for at least 20 minutes with fully charged battery to guarantee that the device would never go into the battery-saving mode. No other apps were active during the experiments. The experiment results show that Device 1 consumed 102 milliampere hour (mAh) energy in 20 minutes, which corresponds to 3.4% of the full battery capacity. On the other hand, Devices 2 and 3 consumed higher amounts of energy as compared to Device 1 (about 136 mAh on Device 2 and 141 mAh on Device 3). 20 minutes was considered an upper bound of what users might use the app continuously in any one setting. In practice, most usage is completed within several minutes.

Furthermore, it is important to ensure that the mobile app's performance does not degrade and that it runs smoothly on different vendors' devices with different specifications. Figure 12 (b) depicts the result of the mobile app's performance analysis in which response time of five different functions is measured on each device. The results illustrate that response time was slightly increased on Devices 2 and 3 as compared to Device 1, but they still output the correct response within 1500 ms maximum.

VII. CONCLUSION

This paper presented an aspect-based sentiment classification framework that classifies opinions/reviews about aspects into positive or negative. In this framework, a tree-based aspects extraction method is proposed that extracts both explicit and implicit aspects from tourist opinions. It extracts frequent nouns and noun phrases from reviews text, and then groups similar nouns using WordNet. Decision tree is employed on reviews where review words are used as internal nodes and extracted noun as leaf of a tree. Opinionless and irrelevant sentences are first removed by employing Stanford Basic Dependency on each sentence. Next, features are extracted from the remaining sentences with N-Grams and POS Tags to train the classifiers. Lastly, machine learning algorithms are applied to the extracted features to train the classifiers. Once suitably trained, the learned model is used to classify the sentiment about extracted aspects into positive or negative.

Experiments have been conducted on real-world datasets taken from restaurant and hotel reviews websites and social media platforms to evaluate the proposed framework. Comparative evaluations conducted showed that the proposed framework outperformed existing methods used in the tourism domain. Specifically, in the task of identification of aspects, the proposed method correctly identified 80% of the relevant aspects in the restaurants dataset and 85% in the hotels dataset. In the aspect-based sentiment classification task, the best results among the five machine learning algorithms evaluated were achieved with NBM. It correctly classified 88.08% of the aspects on the restaurants dataset and achieved 90.53% accuracy on and hotels dataset.

Finally, the proposed aspect-based sentiment mining framework has been implemented as a mobile app for consumers. The mobile app provides a user-friendly way to analyze tourist sentiments by aspects. Using their smartphones, traveling consumers can get useful and noise suppressed information to help them make decisions when they visit any tourist place. Future research will focus on scalability and speeding up the total response time to further improve the user experience.

REFERENCES

[1] Y. Blanco-Fernandez, M. Lopez-Nores, J. J. Pazos-Arias, A. Gil-Solla, and M. Ramos-Cabrer, "Exploiting digital TV users' preferences in a tourism recommender system based on semantic reasoning," *IEEE Trans. Consum. Electron.*, vol. 56, no. 2, pp. 904–912, May 2010. doi: 10.1109/TCE.2010.5506018.

[2] C. S. Khoo and S. B. Johnkhan, "Lexicon-based sentiment analysis: Comparative evaluation of six sentiment lexicons," *J. Inf. Sci.*, vol. 44, no. 4, pp. 491–511, Aug. 2018. doi: 10.1177/0165551517703514.

[3] R. L. Rosa, D. Z. Rodriguez, and G. Bressan, "Music recommendation system based on user's sentiments extracted from social networks," *IEEE Trans. Consum. Electron.*, vol. 61, no. 3, pp. 359–367, Aug. 2015. doi: 10.1109/TCE.2015.7298296.

[4] R. Moraes, J. F. Valiati, and W. P. G. Neto, "Document-level sentiment classification: An empirical comparison between SVM and ANN," *Expert Syst. Appl.*, vol. 40, no. 2, pp. 621–633, Feb. 2013. doi: 10.1016/j.eswa.2012.07.059.

[5] A. Moreo, M. Romero, J. L. Castro, and J. M. Zurita, "Lexicon-based comments-oriented news sentiment analyzer system," *Expert Syst. Appl.*, vol. 39, no. 10, pp. 9166–9180, Aug. 2012. doi: 10.1016/j.eswa.2012.02.057.

[6] G. Wang, J. Sun, J. Ma, K. Xu, and J. Gu, "Sentiment classification: The contribution of ensemble learning," *Decis. Support Syst.*, vol. 57, pp. 77–93, Jan. 2014. doi: 10.1016/j.dss.2013.08.002.

[7] J. C. De Albornoz, L. Plaza, P. Gervás, and A. Díaz, "A joint model of feature mining and sentiment analysis for product review rating," in*Proc. Eur. Conf. Inf. Retrieval*, 2011, pp. 55–66. doi: 10.1007/978-3-642-20161-5_8.

[8] E. Marrese-Taylor, J. D. Velásquez, and F. Bravo-Marquez, "Opinion zoom: A modular tool to explore tourism opinions on the Web," in *Proc. IEEE/WIC/ACM Int. Joint Conf. Web Intell. Intell. Agent Technol.*, 2013, pp. 261–264. doi: 10.1109/WI-IAT.2013.193.

[9] E. Marrese-Taylor, J. D. Velásquez, and F. Bravo-Marquez, "A novel deterministic approach for aspect-based opinion mining in tourism products reviews," *Expert Syst. Appl.*, vol. 41, no. 17, pp. 7764–7775, Dec. 2014. doi: 10.1016/j.eswa.2014.05.045.

[10] Z. Hai, K. Chang, J.-J. Kim, and C. C. Yang, "Identifying features in opinion mining via intrinsic and extrinsic domain relevance," *IEEE Trans. Knowl. Data Eng.*, vol. 26, no. 3, pp. 623–634, Mar. 2014. doi: 10.1109/TKDE.2013.26.

[11] M. Afzaal, M. Usman, A. C. M. Fong, S. Fong, and Y. Zhuang, "Fuzzy aspect based opinion classification system for mining tourist reviews," *Adv. Fuzzy Syst.*, vol. 2016, Oct. 2016, Art. no. 2. doi: 10.1155/2016/6965725.

[12] M. Colhon, C. Bădică, and A. Şendre, "Relating the opinion holder and the review accuracy in sentiment analysis of tourist reviews," in *Proc. Int. Conf. Knowl. Sci. Eng. Manag.*, 2014, pp. 246–257. doi: 10.1007/978-3-319-12096-6_22.

[13] A. Mukherjee and B. Liu, "Aspect extraction through semi-supervised modeling," in *Proc. 50th Annu. Meeting Assoc. Comput. Linguist.*, 2012, pp. 339–348.

[14] H. Wang, Y. Lu, and C. Zhai, "Latent aspect rating analysis on review text data: A rating regression approach," in *Proc. 16th ACM SIGKDD Int. Conf. Knowl. Disc. Data Min.*, Washington, DC, USA, 2010, pp. 783–792. doi: 10.1145/1835804.1835903.

[15] J. Zhu, H. Wang, M. Zhu, B. K. Tsou, and M. Ma, "Aspect-based opinion polling from customer reviews," *IEEE Trans. Affect. Comput.*, vol. 2, no. 1, pp. 37–49, Jan./Jun. 2011. doi: 10.1109/T-AFFC.2011.2.

[16] Y. Wu and M. Ester, "Flame: A probabilistic model combining aspect based opinion mining and collaborative filtering," in *Proc. 8th ACM Int. Conf. Web Search Data Min.*, 2015, pp. 199–208. doi: 10.1145/2684822.2685291.

[17] F. Xianghua, L. Guo, G. Yanyan, and W. Zhiqiang, "Multi-aspect sentiment analysis for Chinese online social reviews based on topic modeling and HowNet lexicon," *Knowl. Based Syst.*, vol. 37, pp. 186–195, Jan. 2013. doi: 10.1016/j.knosys.2012.08.003.

[18] X. Xu, X. Cheng, S. Tan, Y. Liu, and H. Shen, "Aspect-level opinion mining of online customer reviews," *China Commun.*, vol. 10, no. 3, pp. 25–41, Mar. 2013. doi: 10.1109/CC.2013.6488828.

[19] M. Pontiki *et al.*, "SemEval-2016 task 5: Aspect based sentiment analysis," in *Proc. 10th Int. Workshop Semantic Eval.*, 2016, pp. 19–30.

[20] M. Pontiki, D. Galanis, H. Papageorgiou, S. Manandhar, and I. Androutsopoulos, "SemEval-2015 task 12: Aspect based sentiment analysis," in *Proc. 9th Int. Workshop Semantic Eval.*, 2015, pp. 486–495.

[21] L. Zhang and B. Liu, "Aspect and entity extraction for opinion mining," in *Data Mining and Knowledge Discovery for Big Data*. Berlin, Germany: Springer, 2014, pp. 1–40. doi: 10.1007/978-3-642-40837-3_1.

[22] S. Schuster and C. D. Manning, "Enhanced English universal dependencies: An improved representation for natural language understanding tasks," in *Proc. LREC*, May 2016, pp. 23–28.

[23] Y. Liu and S. Liao, "Preventing over-fitting of cross-validation with kernel stability," in *Proc. Joint Eur. Conf. Mach. Learn. Knowl. Disc. Databases*, 2014, pp. 290–305. doi: 10.1007/978-3-662-44851-9_19.

[24] A. M. Kibriya, E. Frank, B. Pfahringer, and G. Holmes, "Multinomial naive Bayes for text categorization revisited," in *Proc. Aust. Joint Conf. Artif. Intell.*, 2004, pp. 488–499. doi: 10.1007/978-3-540-30549-1_43.

[25] J. A. Suykens and J. Vandewalle, "Least squares support vector machine classifiers," *Neural Process. Lett.*, vol. 9, no. 3, pp. 293–300, Jun. 1999. doi: 10.1023/A:1018628609742.

[26] A. L. Berger, V. J. D. Pietra, and S. A. D. Pietra, "A maximum entropy approach to natural language processing," *Comput. Linguist.*, vol. 22, no. 1, pp. 39–71, Mar. 1996. doi: 10.1.1.103.7637.

[27] P. Jiang *et al.*, "RF-DYMHC: Detecting the yeast meiotic recombination hotspots and coldspots by random forest model using gapped dinucleotide composition features," *Nucleic Acids Res.*, vol. 35, no. S2, pp. W47–W51, Jul. 2007. doi: 10.1093/nar/gkm217.

[28] I. N. Athanasiadis, V. G. Kaburlasos, P. A. Mitkas, and V. Petridis, "Applying machine learning techniques on air quality data for real-time decision support," in *Proc. 1st Int. NAISO Symp. Inf. Technol. Environ. Eng.*, Gdansk, Poland, Jun. 2003, pp. 51–56.

[29] K. Toutanova, D. Klein, C. D. Manning, and Y. Singer, "Feature-rich part-of-speech tagging with a cyclic dependency network," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguist. Human Lang. Technol.*, 2003, pp. 173–180. doi: 10.3115/1073445.1073478.

[30] G. A. Miller, "WordNet: A lexical database for English," *Commun. ACM*, vol. 38, no. 11, pp. 39–41, Nov. 1995. doi: 10.1145/219717.219748.

[31] V. Svetnik *et al.*, "Random forest: A classification and regression tool for compound classification and QSAR modeling," *J. Chem. Inf. Comput. Sci.*, vol. 43, no. 6, pp. 1947–1958, Nov./Dec. 2003. doi: 10.1021/ci034160g.

[32] W. Oliveira, R. Oliveira, and F. Castor, "A study on the energy consumption of Android app development approaches," in *Proc. 14th IEEE/ACM Int. Conf. Min. Softw. Repositories*, 2017, pp. 42–52. doi: 10.1109/MSR.2017.66.

[33] K. A. Phan, Z. Tari, and P. Bertok, "A benchmark on SOAP's transport protocols performance for mobile applications," in *Proc. ACM Symp. Appl. Comput.*, 2006, pp. 1139–1144. doi: 10.1145/1141277.1141548.

Muhammad Afzaal received the B.S.C.S. degree in computer sciences from MI University, Azad Kashmir, Pakistan, in 2012, and the M.S. degree in computer sciences from the Shaheed Zulfikar Ali Bhutto Institute of Science and Technology, Islamabad, Pakistan. He has published in international journals and conference proceedings. His main areas of interest are data mining, machine learning, and sentiment analysis.

Muhammad Usman received the Ph.D. degree in computer and information sciences from the Auckland University of Technology, New Zealand. He is currently with the Shaheed Zulfikar Ali Bhutto Institute of Science and Technology, Islamabad, Pakistan. He has published in international journals and conference proceedings. His research interests include data mining and machine learning. He has served as a reviewer for a number of premier journals and conferences.

Alvis Fong (SM'04) received the four degrees in EE and CS. He was a Professor of computer science with the Auckland University of Technology. He is currently with Western Michigan University. His research interests include intelligent systems, information retrieval, and multimedia. He has published circa 190 papers in the above areas.