



# Implicit Aspect-Based Opinion Mining and Analysis of Airline Industry Based on User-Generated Reviews

Kanishk Verma<sup>1</sup> · Brian Davis<sup>1</sup>

Received: 18 November 2020 / Accepted: 27 April 2021 / Published online: 21 May 2021  
© The Author(s) 2021

## Abstract

Mining opinions from reviews has been a field of ever-growing research. These include mining opinions on document level, sentence level and even aspect level. While explicitly mentioned aspects from user-generated texts have been widely researched, very little work has been done in gathering opinions on aspects that are *implied* and not explicitly mentioned. Previous work to identify implicit aspects and opinion was limited to syntactic-based classifiers or other machine learning methods trained on restaurant dataset. In this paper, the present is a novel study for extracting and analysing implicit aspects and opinions from airline reviews in English. Through this study, an airline domain-specific aspect-based annotated corpus, and a novel two-way technique that first augments pre-trained word embeddings for sequential with stochastic gradient descent optimized *conditional random fields* (CRF) and second using machine and ensemble learning algorithms to classify the implied aspects is devised and developed. This two-way technique resolves double-implicit problem, most encountered by previous work in implicit aspect and opinion text mining. Experiments with a hold-out test set on the first level i.e., entity extraction by optimized CRF yield a result of ROC-AUC score of 96% and  $F_1$  score of 94% outperforming few baseline systems. Further experiments with a range of machine and ensemble learning classifier algorithms to classify implied aspects and opinions for each entity yields a result of ROC-AUC score ranging from 71 to 94.8% for all implied entities. This two-level technique for implicit aspect extraction and classification outperforms many baseline systems in this domain.

**Keywords** Conditional random field · Stochastic gradient descent · Machine learning · Ensemble learning · Implicit aspects · Augmenting word embeddings · Classification · Corpus · Sequential labelling

## Introduction

Travel and tourism are well-liked terms amongst all generations of people. The airline industry is a key facilitator in this domain. For this industry, serving its customers with not only cost-effective but also satisfactory service options is paramount [1]. Opinions are very important to businesses and organizations, because they always want to

find customer or public opinions about their products and features [2]. In this 21st information age, with constant development in social and web media, a multitude of platforms are available like *Trip Advisor*, *Airline Ratings* etc., for consumers to express their views on air travel. This serves in favour of the airline companies, as it becomes their one-stop to access rich customer feedback information. However, many times, due to a variety of reasons like paid promotions, fraudulent and unstructured nature of these reviews, insightful information cannot be extracted. Therefore, a need is felt to have a mechanism that gathers cognizance in terms of the perception of customers on airline-specific aspects [3].

Lui and Zhang et al. defined the term opinion as “*a concept covering sentiment, evaluation, appraisal or attitude held by a person*” [2]. Aspects and entities are more like topics in a text document. Hu and Liu et al. coined this type of analysis as feature-based sentiment analysis. [4] Aspect or entity-based analysis identifies the target of the opinion. It is a fine-grained approach to text analysis.

---

This article is part of the topical collection “Intelligent Computing and Networking” guest edited by Sangeeta Vhatkar, Seyedali Mirjalili, Jeril Kuriakose, P.D. Nemade, Arvind W. Kiwelekare, Ashok Sharma and Godson Dsilva.

---

✉ Kanishk Verma  
kanishk.verma@adaptcentre.ie

Brian Davis  
brian.davis@adaptcentre.ie

<sup>1</sup> ADAPT SFI Centre, School of Computing, Dublin City University, Dublin, Ireland

## Paper Nomenclature

In this paper, an *entity* is the feature of the airline and an *implicit aspect* or sub-aspect is its *attribute*. Examples for entities are *food*, *cabin*, *seat*, *staff*, etc. These entities within themselves have various attributes associated with them, it becomes important to divide them, it becomes important to divide them further into sub-aspect or implicit aspects.

For example, a sentence in a review could read “*the cabin was cold, smelly and a bit weary*”. Here, the entity *cabin* is accompanied its attributes like *temperature*, *fragrance*, and *condition*. The phrases or terms like “*cold*”, “*smelly*”, and “*a bit weary*” are terms that imply an opinion to each individual attribute of the entity *cabin*. This paper devises a technique to identify **airline-specific** entities from such implicit phrases or terms. This approach helps in making a fine-grained analysis of opinions and maps them accurately to respective entity-implicit aspect-pair.

## Research Motivation

Understanding which passenger airline industry-specific aspects can be leveraged for implicit aspect-based opinion mining is one of the key focus of research. In addition, we will develop novel domain specific opinionated corpus annotated with implicit aspects. Furthermore, we experiment with specific lexicon<sup>1</sup> generation techniques for influencing this type of opinion mining.

## Data

Trip advisor and Airline ratings are online microblogging platforms primarily used of viewing reviews and experiences of travellers either travelling, to the same destination or other, all over the globe. Usually, people before making airline ticket purchases do read reviews [4].

In this study, 3000 reviews were collected within a period of 1 month with an aim to study public opinion with respect to 16 Airlines (see online Appendix A). From these 3000 reviews, after curating, only 1803 reviews were determined to be relevant for this study. Detailed statistical analysis was carried on to understand the quality of it. This statistical analysis information is available in Table 1.

In summary, the goal of this study is extract implied aspects and opinions from airline reviews. To achieve this goal, a new dataset was created, which to our knowledge, is the first time a dataset specifically for implicit aspects of airline reviews is created. Using a *supervised lexicon-based*

<sup>1</sup> Lexicon: It is a component of natural language processing that contains grammatical information about individual words or strings.

**Table 1** Dataset statistics

Number of reviews	1803
Number of sentences	9591
Average number of sentences in a review	5
Type token ratio [5]	0.27
Most common word	Flight
Average word length	7
Labelled unique word corpus size	3280

**Table 2** Entity-wise implicit aspect list

Entity	Implicit aspect(s)
Food	<i>Service</i> <i>Temperature</i> <i>Taste</i>
Entertainment	<i>Visual</i> <i>Audio</i> <i>General</i>
Cabin	<i>Condition</i> <i>Fragrance</i> <i>Size</i> <i>Temperature</i>
In-flight service	<i>Operations</i> <i>Facility</i>
Off-flight service	<i>Ticketing</i> <i>General</i> <i>Facility</i>
Staff	<i>Behaviour</i> <i>General</i>
Seat	<i>Operations</i> <i>Comfort</i>
Possession	<i>Handling</i> <i>General</i>

*technique*, few experiments were run to gather insightful information about airline-based implied aspects and opinions. The results of which were favourable for the study. Furthermore, in this paper, discussions are on methodology, issues and challenges, experimental setup, and evaluations/results of this approach (Tables 2, 3, 4, 5).

## Methodology

The methodology of this study consists of multiple modules. Each module was developed keeping in mind that the dataset is fresh, new and one of kind. Therefore, the methodology pipeline includes *data collection*, *corpus statistics*, *annotation*, *feature engineering*, *sequence labelling*, and *classification tasks*.

**Table 3** Feature engineering tasks

Word features	Parts-of-speech tags; Dependency parsing
Numeric word representation	Count vectorizer; term frequency— inverse document frequency; Augmenting word embeddings

**Table 4** Type token ratio scores

Source	TTR score
Trip Advisor	0.35
Airline ratings	0.37

**Table 5** Detailed example of Level-1 annotation

<i>INPUT:</i> “Overall the experience was comfortable and spacious with delicious meals”
<i>Output:</i> [(“experience was comfortable”, “Inflight”), (“spacious”, “cabin”), (“delicious meals”, “food”)]

### Entity and Aspect Selection

Post-dataset statistical analysis, the two annotators carefully read about 500 reviews. Features of the passenger aircraft, services offered by the airlines both in and off the flight were formulated in a list. After curating the list, a data-driven decision led to enlist entities into eight categories. The representation of these eight implicit entity-aspect pairs can be found in Table 2.

### Data Annotation

Manual annotation and labelling of all the reviews were done using Doccano [6] annotation tool. An inter-annotator agreement guideline [7] was also set up. (See online appendix A). Annotation was done on two levels i.e., *entity level* and *implicit aspect level*. Cohen’s Kappa coefficient [8] was chosen to find quality of annotation by annotators. The results of which are shared in Evaluation section.

### Feature Engineering

The feature engineering task was divided in two as described in Table 3, one to capture word features and the other to gather numeric representations of the word features. (See Appendix B).

### Augmenting Word Embeddings

The numeric representations like count vectorizer and TF-IDF are frequency based and lack contextual information. [9] Due to the limited size of the dataset, a need was felt to augment<sup>2</sup> pre-trained word embeddings. Pre-trained Glove [10] vectors trained on user-generated text was used. These pre-trained vectors were augmented with Word2Vec [11, 12] for corpus embeddings. Also, the parameters augmented were the one’s that considered maximum distance between focus word and its contextual neighbours (See online Appendix D).

### Sequence Labelling with Conditional Random Fields

Sequence Labelling is a supervised learning<sup>3</sup> task where a label is assigned to each element of a sequence. For our study, to extract words and classify them into respective entities, a conditional random field algorithm was selected. Conditional random fields [13] adjust to a variety of statistically correlated features as input just like a sequential classifier. Also, like a generative probabilistic model, it trades-off decision at different sequence to obtain a global optimal labelling. (See online Appendix E).

The CRF model was optimized using stochastic gradient descent<sup>4</sup> with L2 regularization.<sup>5</sup> This is done to maximise the likelihood of the CRF and can be represented as follows:

$$\log P(Y|X) = w \cdot \varphi(Y, X) - \left( \log \sum_{Y^T} e^{w \varphi(Y, X)} \right) \tag{1}$$

After taking derivatives on the above equation, we get below

$$\frac{d}{dw} \log P(Y|X) = \varphi(Y, X) - L^2 \sum_{Y^T} P(Y^T|X) \varphi(Y^T, X), \tag{2}$$

where it means  $\varphi(Y, X)$  add correct features and subtract  $P(Y^T|X)$  which is expectation of features and L2 is a regularization penalty term.

<sup>2</sup> Made the word embeddings larger and stronger by adding Gloving embeddings.

<sup>3</sup> It means learning a mapping between a set of input variables X and output variables Y and applying these mappings to make predictions from unseen data.

<sup>4</sup> Gradient Descent: An optimization algorithm used to minimize some function iteratively.

<sup>5</sup> L2 regularization: It is a penalty regularization technique which does not let the algorithm over-fit.

**Table 6** Detailed example of Level-2 annotation

*INPUT:* [(“experience was comfortable”, Inflight), (“spacious”, cabin), (“delicious meals”, “food”)]

*OUTPUT:* [[(“experience”, inflight- operations), (“comfortable”, inflight- operations)], [(“spacious”, cabin-size)], [(“delicious”, food-taste), (“meals”, food-service)]]

### Classification for Implicit Aspect Extraction

The aspect extraction task needed classifier models that could accurately predict the aspect. Different algorithms were used to classify and compare how accurate each model was to classify these sub-aspects. Algorithms like Support Vector Machine, Decision Trees, Random Forest, a bagging ensemble learning algorithm Voting Classifier and a boosting ensemble learning algorithm XGBOOST were employed. (See online Appendix F).

### Data Setup

#### Data Pre-Processing

Using standard pre-processing techniques like removing domain-specific stop words, removal of unnecessary punctuations, spell correction, converting numbers to words, and word standardization. Since, the data were user-generated, there were many contractions of words, for example, “couldn’t”, “can’t”, “aren’t” etc., were seen quite often in the texts. Therefore, fixing these contraction words was also a part of the study. The contraction words were replaced with their expanded words (See online Appendix G).

#### Corpus Statistics

The data being user-generated were raw and unstructured. It is the first this group of reviews was considered for text mining and analyzing. Therefore, two statistical strategies, viz, type-token ration [5] and Zipf’s distribution [14] were used to determine variability in the dataset.

Type Token Ratio (TTR) is represented as follows (See online Appendix H):

$$\text{TTR} = \frac{(\text{number of types})}{(\text{number of tokens})}. \quad (3)$$

TTR Scores are low for both data sources as seen in Table 4, this means that there are many repeated terms in the corpus. (See online Appendix H).

Zipf’s law states that a relationship between frequency of word ( $f$ ) and its position in the list i.e., its rank ( $r$ ) is inversely proportional to one another

$$f \propto \frac{1}{r} \quad (4)$$

### Manual Annotation

As explained in the methodology, the annotation was done on two levels using Doccano. There are detailed examples and explanation of this manual annotation strategy.

Once, entity-level tuples<sup>6</sup> were tagged containing a word or word phrases with entity-name, as seen in Table 5. After completing entity-level annotation, another fine-grained approach to classify entity-wise word or word phrases to their respective implied aspects was conducted, details of which are available in Table 6.

### Inter-annotator Agreement

As explained in the methodology of this experimental study, after adhering with the guidelines in the inter-annotator agreement, and using Python’s sk-learn Kappa score library, the Cohen’s Kappa [8] score for agreement level of annotators was calculated (Tables 7, 8, 9).

### Experimental Setup

#### Training Data Preparation

This experiment study used techniques described in the methodology section for preparing the training data. Taking an example sentence, this process will be explained in detail. Example sentence: “Overall, the experience was comfortable and spacious with delicious meals”. Table 7 denotes entity-level and implicit aspect-level annotations for the example sentence. From this review, words like experience, comfortable, spacious, delicious, and meals were identified as aspect terms and their semantic and syntactic information was extracted by parsing them through off-the-shelf state-of-the-art models like Stanford Core NLP API [15] to extract part-of-speech (POS) tags and

<sup>6</sup> Tuples are a data type that is similar but also distinct from the list data type. The instance are characterized by having fixed attributes and the elements of tuple instance can differ to another in data type.

**Table 7** Annotated and labelled list of example sentence

Entity level	
Entity	Word/phrases
In-flight service	Experience was comfortable
Cabin	Spacious
Food	Delicious meals
Implicit level	
Aspect	Word
Inflight operations	Experience
Inflight operations	Comfortable
Cabin size	Spacious
Food taste	Delicious
Food service	Meals

**Table 8** Entity-ID list

Entity	Entity-ID
Food	f
Cabin	c
Entertainment	e
Staff	st
Seat	s
Off-flight	o
In-flight	i
Possession	p

**Table 9** TF-IDF vectorization

Word	TF-IDF score
Basic	0.965545
Redemption	0.965545
Rescue	0.958253

dependency tags and gathering their sentiment using Vader (See online Appendix B).

Using these techniques, a list of features was generated which consisted of main-word, main-word POS tag, dependent word, dependent word POS tag, main-word sentiment score, dependent word sentiment score, previous and next word.

For the task of sequence labelling to identify the entity, a word or word phrase belongs to, the tuples were added with their respective labels i.e., the label added to a tuple was the label the “main word” belonged to.

For example, a Tuple: (“delicious”, “JJ”, “meals”, “NNS”, 0.6, 0.0, “advmod”, “spacious”, “meals”) has the main word *food*, so a new entry to this was made as “f”,

which became the Y or the dependent variable. After getting results from the CRF model, the entity-id i.e., it was classified as “*food*”.

Once the correct entity is identified, the next step is to classify which aspect is mentioned in the sentence. Later, the Entity-ID is added as seen in Table 8 to the training data and then vectorized.

### Count Vectorization

For this experiment study, since the methodology does try to keep certain punctuations and special characters, there is a need to create its own vectorizer.

The results for an example sentence are as follows,

Sentence: “so overall I highly recommend this airline!”.

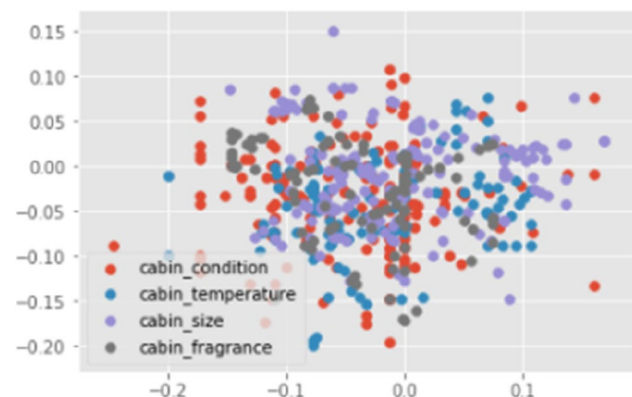
Vector: {“so”:6, “overall”:4, “I”:3, “highly”:1, “recommend”:5, “this”: 7, “airline”:0, “!”:2}.

### TF-IDF Vectorization

For this experiment study, the TF-IDF score for the words in the feature sets was calculated using python’s sci-kit TF-IDF vectorized. Table 9 shows the result of TF-IDF for few corpus words.

### Augmenting Word Embeddings

As mentioned earlier, a word embedding model using Word2Vec for this corpus was trained. And a pre-trained Twitter Glove Embeddings consisting of vocabulary size of 1.2 million words and 27 billion tokenized twitter words with a 100-dimensional vector was chosen.



**Fig. 1** Cabin class imbalance rectified with SMOTE

Using the algorithm 1, a new set of vector embeddings were merged with the pre-trained Glove embeddings.

**Algorithm 1** Word embedding vector generation using pre-trained glove vectors

---

**Inputs**  
 $S = [W_1, W_2, W_3, \dots, W_n]$ , Input sentence  $S$  contains  $n$  words  
 $path$  = path where downloaded embeddings are stored  
 $GloveVec$  = Pretrained Glove Vectors

**Output**  
 $Word2Vec$  Embedding Model

---

```

// Load the Glove Pre-trained Vectors
with open(path):
    gloveVec = embedding vectors // Create Word2Vec Embedding for Airline Corpus
word2vec = Word2Vec Create Model

for word, vector in zip(word2vec.index2word, word2vec.vectors) do
    | w2v = dict(word: vector)
end

// Vectors for airline Corpus are loaded
for each  $W_i$  in  $S$  do
    if  $W_i$  exists in  $gloveVec$  then
        | extract  $vecW_i$   $MV_i = vecW_i$ end
    else if  $W_i$  exists in  $w2v$  then
        | extract  $vecW_i$   $MV_i = vecW_i$ end
    else
        | extract  $vecW_i$   $MV_i = generateNewvecW_i$ end
end

```

---

With this algorithm 1, a new set of word embeddings were generated to vectorize textual information in the feature tuple.

### Cosine Similarity Index

Along with the word embeddings, cosine similarity between main and dependent word was added as a new feature. (See online Appendix D).

These new features were then used to classify opinionated texts into their respective implicit-aspect classes.

### Handling Class Imbalance

After annotation, there was a high imbalance amongst implicit aspect classes of almost all entities. This imbalance was handled using an oversampling technique called Synthetic Minority Oversampling Techniques [16] (See online Appendix F). SMOTE was performed for all eight entities.

Results of SMOTE imbalance handling for entity: *Cabin* is as follows:

Class: {"Condition":182, "Size": 182, "Temperature":117, "Fragrance": 102}.

This could be visualized as a scatter distribution show in Fig. 1 below.

**Table 10** ROC-AUC scores for classification of entities

Entity	Algorithms				
	S (%)	D (%)	R (%)	V (%)	X (%)
Food	84	92	94	94.8	94.7
Cabin	75	75	85	85.6	77
Entertainment	73.6	79.9	83.1	84.3	85.9
In-flight	60.3	70.3	72.2	74.9	71.2
Off-flight	66.4	86.2	84.9	84.8	89.8
Possession	66.9	66.9	70.5	73.3	73.4
Seat	66	73.7	75	75.7	78
Staff	75.6	76.9	80.9	82.1	81.4

### Implicit Aspect Classification

A total of eight models were created for each entity i.e., there are independent classification models for training to classify each entity. The reason for creating eight models is to devise a perfect a model for recognizing and classifying each *Entity* with its own *Implicit Aspect*.

This experiment study makes use of state-of-the-art classification algorithms. Three of which were ensemble learning techniques. These include Gradient boosting algorithm—XGBOOST, a Voting Bagging algorithm using three tree-based classification techniques Decision Trees, Random Forest, and Extra Trees Classifier. And other machine learning algorithms like SVM, Decision Tree.

The reason for using these different algorithms was to gather insightful information on the performance of classification which was evaluated based on ROC-AUC [17] score and  $F_1$  [18] scores. (See online Appendix I).

### Evaluation and Results

This experimental study using state-of-the-art techniques and algorithms is a new approach to mine and extract implicit aspects from opinionated texts. The first evaluation was for the annotation of the dataset using Cohen's Kappa Co-efficient. The two annotators agreement scores ranged from 80.48 to 82.13% for entity level and implicit aspect level annotation (See online Appendix A).

The impact of using this novel two-level technique while annotation and training for classification help overcome the double-implicit problem. The decision to augment pre-trained word embeddings has been beneficial to build a contextually powerful embedding model. Put-together this empowers the ensemble learning classification algorithms to provide better classification results, which is observed through the ROC-AUC and F-statistic scores.

The second evaluation was for the sequence labelling task using stochastic gradient descent with L2 regularization Conditional Random Field. This was to classify texts in eight different entities.

The ROC-AUC score achieved for this task is 96.5% and  $F_1$  score of 94.56% (See online Appendix I).

The third evaluation was for the classification task using five different classification algorithms. A detailed ROC-AUC score evaluation metric is available in Table 10 (Highlighted in green provides best score) (For further details, see online Appendix I).

In the above table, S stands for Support Vector Machines, D for Decision Trees, R for Random Forest, V for Voting Classifier, and X for XGBOOST algorithms. In all these machine learning and ensemble learning classification algorithms, the bagging technique outperformed all other classification algorithms (See online Appendix I).

## Issues and Challenges

Manual annotation was a big challenge. Everyone has a different outlook on implied meanings. One can think of words like “boarding, de-boarding, take-off” as in-flight operations. But, if someone spends a little time to go through the review, one can understand the concept terms “boarding, de-boarding, take-off” are off-flight facilities provided by the airlines. Therefore, using corpus statistics techniques and adhering to the inter-annotator guidelines, the annotators made mutually agreeable decisions (See online Appendix A).

The word *spacious* in the dataset was challenging for the labelling task. It is a word that was frequent in the reviews. Also, if used within the same sentence or context of “cabin”, it means that the “cabin” was “big” implying to the size of the cabin. But in the context of “seat”, it implies that the “seat” had ample leg room implying “comfort”. This word has two implicit meanings thereby formed a double-implicit problem. Such a problem was tackled using T-distributed stochastic nearest neighbours for word embeddings dimension reduction and clustering technique [19]. This allowed word distances of these double-implicit words to be mapped with each implicit aspect-entity pair. Wherever the words were close, it was mapped to the respective implicit aspect-entity pair. (See online Appendix D).

For example, “spacious” occurs in the same vector space as of “size” for cabin and “comfort” for seat. Therefore, the word cosine distance between spacious, size and comfort were included as a feature.

## Related Work and Improvements

Our research concentrates on implicit aspect extraction, opinion lexicon generation, and engineering an annotated implicit aspect-based sentiment corpus that can influence implicit opinion mining from consumer reviews in the airline industry. Few studies that are done in this realm for implicit aspect-based opinion mining and extraction but very few on implicit aspect-based opinion mining.

In a research study proposed by Chinsha et al. [20], the methodology proposed a syntactic-based approach using dependency parsing, and another research for comparing word representations for implicit classification [21], make use of SentiWorNet and have dataset restrictions.<sup>7</sup> The present study extends the result of these two papers, using syntactic approach to group implicit aspect synonyms for a larger dataset.

Research dealing with the double-implicit problem in opinion mining and sentiment analysis proposed a protocol to derive a labelled corpus for implicit polarity and aspect analysis [22]. The work in this paper is limited to Chinese restaurant reviews. The present study addresses not only the dataset limitation but also the labelling of the corpus technique using Type/token Ratio and other corpus statistic techniques which are explained in the experimental setup Sect. 4.

Another study using two corpora proposed a hybrid model to support Naïve Bayes training to identify implicit aspects [23]. This corpus and dictionary-based approach is limited to only adjective type words of a sentence. The present study extends this work by taking considering a combination of adjectives, adverbs, nouns, and other part-of-speech indicators and uses ensemble learning for classification.

A study conducted on implicit aspect indicator extraction, models relations between the polarity of a document and its opinion target using Conditional Random Field (CRF) [24]. This method is limited, however, to only cellular device data and the entities are picked from a pre-trained Stanford CRF model. Our work extends Conditional Random Field and extends it to the airline domain.

## Conclusion and Future Work

The present research study using a supervised machine learning approach provides a novel technique to overcome the implicit opinion and aspect mining problem. It does so by, identifying eight different airline industry-specific aspects that can be leveraged for the task of opinion mining. They include fine-grained entities such as the cabin, entertainment, food, in-flight service, off-flight service, seat, staff,

<sup>7</sup> A methodology that is used to extract grammatical structure from sentences.

and possessions. The annotation is done on two levels, one on the entity level and the other is on the sub-aspect level, which allows for a more detailed label construction. The two annotators in this experiment study have a very good agreement on annotated terms. This can be reflected by Cohen's Kappa score ranging from 0.77 to 0.80. Therefore, it can be said that the corpus derived from this study, can be used as a gold standard for implicit aspect-based mining tasks for airline reviews.

This experimental study presents a novel approach of dividing the implicit aspect-based opinion mining task into two levels, one using stochastic gradient descent with L2 regularization for improving conditional random fields to identify entities. This is done with a ROC-AUC Score of **96.58%**, F-statistic score of **94.56%**, and with **0.01** degrees of a mean absolute error on testing data. The second level is to classify each entity into an implicit aspect sub-group. For this state-of-the-art machine and ensemble learning algorithms are used. From the experiments, it is found that ensemble learning outperformed the machine learning approaches. The ROC-AUC scores for ensemble learning algorithms like Voting Classifier range from **73 to 94.8%** and the boosting algorithm like XGBOOST range from **71 to 94.7%** for all eight entities. Synthetic Minority Over-sampling technique proved to be an effective performance improver for the classification and extraction of implicit aspects tasks.

The scope of this experimental study is limited to a few reviews, as possible future work, another study can carry forward the methods proposed in this paper to a larger dataset. Also, another possible future work can be implementing a neural architecture of these proposed methods.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s42979-021-00669-7>.

**Funding** Open Access funding provided by the IReL Consortium. The proposed study was conducted under the guidance and supervision of Dr. Brian Davis at Dublin City University, Ireland for master's thesis. This research was partially conducted with the financial support of Science Foundation Ireland under Grant Agreement No. 13/RC/2106 at the ADAPT SFI Research Centre at Dublin City University.

**Data Availability** The manually annotated corpus data is publicly available under Creative Commons Attribution 4.0 International through Zenodo <https://zenodo.org/record/4126975#.X5RR4lhKjIU>.

## Declarations

**Conflict of Interest** The authors declare that there is no conflict of interest for the presented research study.

**Ethical Approval** The present study has successfully followed all research ethics and post-receiving approval from Dublin City University, Ireland, the study continued. Prior to gathering data from the

websites, written confirmation and approval was collection from TripAdvisor and AirlineRatings.com.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Misopoulos F, Mitic M, Kapoulas A, Karapiperis C. Uncovering customer service experiences with Twitter: the case of airline industry. *Manag Decis.* 2014;52(4):705–23. <https://doi.org/10.1108/MD-03-2012-0235>.
- A Survey of Opinion Mining and Sentiment Analysis | Springer-Link. [https://doi.org/10.1007/978-1-4614-3223-4\\_13](https://doi.org/10.1007/978-1-4614-3223-4_13). Accessed 21 Mar 2021
- Ahn T, Lee T. Service quality in the airline industry: comparison between traditional and low-cost airlines. *Tour Anal.* 2011;16:535–42. <https://doi.org/10.3727/108354211X13202764960582>.
- 'Mining and summarizing customer reviews | Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining'. <https://doi.org/10.1145/1014052.1014073>. Accessed 21 Mar 2021
- Templin MC. *Certain language skills in children; their development and interrelationships*, vol. xviii. Minneapolis: University of Minnesota Press; 1957. p. 183.
- Doccano—Document Annotation Tool. <https://doccano.herokuapp.com/>. Accessed 21 Mar 2021
- The Basics—Natural Language Annotation for Machine Learning [Book]. <https://www.oreilly.com/library/view/natural-language-annotation/9781449332693/ch01.html>. Accessed 21 Mar 2021
- Rosenberg A, Binkowski E. Augmenting the kappa statistic to determine interannotator reliability for multiply labeled data points. In: *Proceedings of HLT-NAACL 2004: Short Papers*, Boston, Massachusetts, USA, May 2004, pp. 77–80, Accessed 21 Mar 2021. Available: <https://www.aclweb.org/anthology/N04-4020>.
- Levy O, Goldberg Y, Dagan I. Improving distributional similarity with lessons learned from word embeddings. *Trans Assoc Comput Linguist.* 2015;3:211–25. [https://doi.org/10.1162/tacl\\_a\\_00134](https://doi.org/10.1162/tacl_a_00134).
- Pennington J, Socher R, Manning C. GloVe: global vectors for word representation. In: *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, Doha, Qatar, Oct. 2014, pp. 1532–43. <https://doi.org/10.3115/v1/D14-1162>.
- Goldberg Y, Levy O. Word2vec explained: deriving Mikolov et al.'s negative-sampling word-embedding method. *ArXiv14023722 Cs Stat.* 2014. Accessed 21 Mar 2021. Available: <http://arxiv.org/abs/1402.3722>.
- Mikolov T, Grave E, Bojanowski P, Puhresch C, Joulin A. Advances in pre-training distributed word representations. *ArXiv171209405*



- Cs, 2017. <http://arxiv.org/abs/1712.09405>. Accessed 21 Mar 2021.
13. Lafferty J, McCallum A, Pereira F. Conditional random fields: probabilistic models for segmenting and labeling sequence data, Dep. Pap. CIS, 2001. Available: [https://repository.upenn.edu/cis\\_papers/159](https://repository.upenn.edu/cis_papers/159).
  14. Powers DMW. Applications and explanations of Zipf's law, 1998. Available: <https://www.aclweb.org/anthology/W98-1218>. Accessed 21 Mar 2021.
  15. Document (Stanford CoreNLP API). <https://nlp.stanford.edu/nlp/javadoc/javanlp-3.5.0/edu/stanford/nlp/dcoref/Document.html>. Accessed 21 Mar 2021.
  16. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. *J Artif Intell Res.* 2002;16:321–57. <https://doi.org/10.1613/jair.953>.
  17. Wu S, Flach P. A scored AUC metric for classifier evaluation and selection. 2005.
  18. Lipton ZC, Elkan C, Narayanaswamy B. Thresholding Classifiers to Maximize F1 Score. *ArXiv14021892 Cs Stat*, May 2014. Available: <http://arxiv.org/abs/1402.1892>. Accessed 21 Mar 2021.
  19. Google News and Leo Tolstoy: Visualizing Word2Vec Word Embeddings using t-SNE | by Sergey Smetanin | Towards Data Science'. <https://towardsdatascience.com/google-news-and-leo-tolstoy-visualizing-word2vec-word-embeddings-with-t-sne-11558d8bd4d>. Accessed 21 Mar 2021.
  20. A syntactic approach for aspect based opinion mining | IEEE Conference Publication | IEEE Xplore'. <https://ieeexplore.ieee.org/abstract/document/7050774>. Accessed 21 Mar 2021.
  21. Braud C, Denis P. Comparing word representations for implicit discourse relation classification. In: Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, Lisbon, Portugal. 2015, pp. 2201–11. <https://doi.org/10.18653/v1/D15-1262>.
  22. Chen H-Y, Chen H-H. Implicit polarity and implicit aspect recognition in opinion mining. In: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), Berlin, Germany. 2016, pp. 20–25. <https://doi.org/10.18653/v1/P16-2004>.
  23. Hybrid approach to extract adjectives for implicit aspect identification in opinion mining | IEEE Conference Publication | IEEE Xplore'. <https://ieeexplore.ieee.org/document/7772284>. Accessed 21 Mar 2021.
  24. Cruz I, Gelbukh A, Sidorov G. Implicit aspect indicator extraction for aspect-based opinion mining, p. 18.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.