Turkish Journal of Computer and Mathematics Education Vol.12 No.13(2021), 2348-2364

Research Article

Ensemble Models for Aspect Category Related ABSA Subtasks

Hetal V[a],. Gandhia,[b], Vahida Z. Attar[c]

[a] Department of Computer Engineering and IT, College of Engineering, Pune, India
[b] Department of Computer Science and Engineering, Walchand College of Engineering, Sangli, India

Article History: Received: 11 January 2021; Revised: 12 February 2021; Accepted: 27 March 2021; Published online: 4 June 2021

Abstract: Many E-commerce and social networking sites have a vast amount of data shared on them. This data is in the form of text, images, audio, and videos. However, people are more accustomed to sharing their experiences or views, about the products purchased by them using textual data. Usually, the users have a good and/or bitter experience about the particular feature of the product, instead of the product as a whole. In this paper, we have performed sentiment analysis of reviews at a deeper level which is known as Aspect Based Sentiment Analysis (ABSA). ABSA allows analyzing data at a finer level. For ABSA, the Aspect Category Detection and Aspect Category Polarity are two subtasks of ABSA related to aspect category. These subtasks aim to detect the aspect categories referenced in the review along with the polarity for each of them. In this paper, we focus on these subtasks for Hindi ABSA Dataset. We compare the different ways of representing the review sentence using word vectors. We compare the performance of the Aspect Category Detection and Aspect Category Polarity subtask using two models. Among the two models- Ensemble model and Feed Forward Neural Network model, the Ensemble model provides significant improvement in performance for both subtasks. The Ensemble model with a sentence vector representation reports considerable improvement in F-score over state-of-the-art Aspect Category Detection results for all four major domains. Our proposed Ensemble model for Aspect Category Polarity subtask provides an increase in accuracy in the range of 7% to 14% for three of the four major domains over best state-of-the-art results.

Keywords: Aspect based sentiment analysis, Ensemble model, Sentence vector, Neural network model, Sentiment Classification

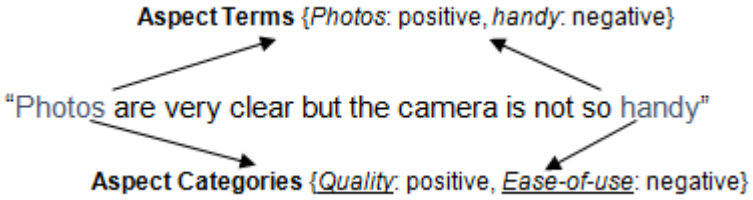
1. Introduction

Social media and e-commerce websites allow people to state their opinions related to an object or entity of interest. Analyzing these opinions at the basic level is possible by performing its document level or sentence level sentiment analysis. Such analysis focuses upon grading the opinion as a whole mostly on a scale of two (*positive*and*negative*) or three (including *neutral*). None of these types of sentiment analysis focus on the features or aspects of the object referenced in it and determines the sentiment linked to those aspects. However, such deeper feature-based opinion summarization [1] is referred to as *Aspect Based Sentiment Analysis* (ABSA) [2]. It allows us to identify the perspective of the opinion holder for each of the different features of the object referenced within that review. For example, for an object say '*camera*', the primary features may include '*Quality*', '*Appearance*', '*Ease-of-use*', '*Price*' etc. With ABSA, it is possible to deduce polarity for each feature as the reviewer might not have the same sentiment towards each feature of the object.

Usually, the Aspect Category is not mentioned explicitly in the review sentence. However, the term Aspect Term, in context to ABSA, refers to some sequence of words that are explicitly mentioned in the review. The sentiment polarity is associated with both Aspect Term and Aspect Category. To give a clear idea, we represent the different terms involved in ABSA with an example review sentence from the '*camera*' domain as given in Figure 1. In the example review sentence, the underlined aspect categories, '*Quality*' and '*Ease-of-use*' have *positive* and *negative* sentiment referenced into it. Thus, each referenced aspect category has its polarity associated with it.

Thus, for ABSA, Aspect Category is the term used to denote the features which are prominent ones for the domain. The set of aspect categories is predefined and predicting the sentiment for each referenced aspect category enables better summarization of opinions. Figure 2 represents the opinion summarization that could be derived when summarizing results by analyzingthe large number of reviews stated for the '*restaurants*' domain. The predefined list of aspect categories is from the '*restaurants*' domain and taken up from a similar analysis done by Ganu et al. [3]. In Figure 2, the x-axis represents each of the aspect categories for the '*restaurants'* domain and such a graph may be represented for each restaurant to compare among the similar features of multiple restaurants. Here, the y-axis denotes the aggregated sentiment grade that can be computed from the regression score for each aspect category.

Figure 1. Similarity and difference between the terms involved in ABSA. The aspect categories are underlined.



The features used for building the models for ABSA subtasks vary over a wider domain including traditional syntactic, lexicon-based, and embedding-based features. Many researchers like Castellucci et al. [4], Brychcin et al. [5] mostly made use of traditional features like Bag-of-words, Bag-of-bigrams for this subtask for English restaurant reviews SemEval 2014 ABSA Dataset. The feature set was extended by seed-oriented features and topic-oriented features [4], ontology features [6] for better performance. The word-vectors for each word in the review sentence [7] and aspect category [8], [9] contribute to semantic features. The word vectors of a review are used for deriving sentence vector or cluster-id [10] and use them as features. The sentiment lexicon-based scores [5], [9] may also be augmented with all these features for building the model.

Many researchers make use of both machine learning and deep learning models [11], [12] for such tasks. However, for ACD subtask, the basis of machine learning models is to address it as a multi-label classification problem. Usually, a separate model is built for each aspect category. The machine learning models like Support Vector Machine [13], [4], [14], and Maximum entropy classifier [10] and [5] are used for this purpose. The deep learning-based approach [8] was proposed for multi-lingual ABSA. The deep learning approaches first derive word embeddings [15] from very large corpora and then build the model. There is extensive use of LSTM, CNN, and their combination [12] and [16] for sentiment classification tasks.

The Ensemble of models with different classifiers or different features with the same kind of classifiers is also proposed for ACP subtask [17], [12]. Ghosh and Sanyal [18] made use of word presence or absence in the document as features and further used the ensemble of multiple feature selection methods. The prominent feature selection was done by the combination of feature selection methods like Information Gain, Gini Index, and Chi-square.

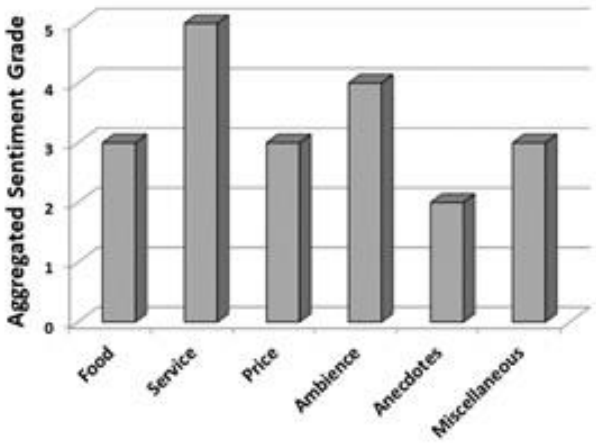
Recently, the sentiment analysis was performed for reviews in different languages like Arabic [19], [20], Czech [10], Hindi [17], etc. Such analysis in multiple languages enables considering the feedback of non-English languages for generating the summary of opinions for a particular product. Thus, customers could check and compare the summary of multiple brands of the same product and decide which brand to purchase.

In this paper, we propose an Ensemble model for Aspect Category Detection (ACD) and ACP subtasks of ABSA of Hindi reviews. The ensemble is a combination of k-Nearest Neighbour (KNN), Support Vector Machine (SVM), and Logistic Regression model. We derive a sentence vector using word-embedding and use it as features for the ACD subtask. The F-score is evaluated and compared with the one obtained using Feed Forward Neural Network (FFNN) based model and also compared with the state-of-the-art results. For the fourth ACP subtask, we extend the feature set by n-gram and Term frequency and inverse document frequency (tf-idf) score.

The major contributions of the paper include:

- We propose different ways of deriving a sentence vector from word embedding vectors and evaluate the performance of the ACD subtask.
- For the ACD subtask of Hindi reviews, our proposed Ensemble-based Model provides a significant performance improvement.
- We make use of a combination of word embedding-based features and lexical features like unigram and bigram features with Term frequency and inverse document frequency (tfidf) score for building the ACP models. In addition to this, we make use of aspect category-based features from the review sentence.
- We have built the models for ACP subtask using the Ensemble approach and report the best performance F-score for three domains than state-of-the-art results.

Figure 2. Opinion summarization of Restaurant reviews on pre-defined aspect categories



The remaining part of the paper is organized as follows:In section 2, we put forth the research work done concerning targeted ACD and ACP subtasks of ABSA. Section 3 describes the different sentence vector representations and other features extracted for classification, along with the models proposed for these two subtasks. In section 4, we analyze the ABSA dataset and share the experiments done and results obtained. Section 5 concludes our work.

2. Related Work

Pontiki et al. [2] coined the term '*aspect category*', for the first time as a part of the ABSA task of English reviews. They released benchmark datasets for the '*laptops*' and '*restaurants*' domains. In this paper, they explained the four subtasks of ABSA in a very simple manner. Here, we first discuss the type of features used and models built for the two ACD and ACP subtasks.

2.1.Aspect Category Detection

The approach used for ACD is mostly supervised and require an ample amount of domain-specific annotated dataset. The supervised approach relies on retrieving hand-crafted features and/ or word embedding-based features to build the model. The models used are Machine Learning based or Deep Learning based. Ganu et al.[3] used the term '*sentence category*', and performed sentence-level analysis by building the SVM model for each category. Few of the different models used by researchers for this subtask of SemEval2014 ABSA Task 4 include- Logistic Regression- Brun et al. [21]; Maximum entropy classifier- Brychcín[5], Hercig et al. [10]; SVM- Kiritchenko et al. [13], Castellucci et al. [4]; Conditional Random Field (CRF)- Patra et al. [22]. Apart from this, the unsupervised approach proposed by Schouten et al. [23] employed the dependency tree-based features (as English ABSA Dataset) for ACD subtask.

Kumar et al. [24] built SVM models for each category with features like Bag-of-words (BOW), Term Frequency (TF) and Inverse Document Frequency (IDF) score, Distributional Thesaurus based features, and Domain Dependency Graph-based features. Lopez et al. [25] also used linear SVM classifiers with their output operated further with category-specific word lists for English and Spanish restaurant review sentences. For Arabic reviews, Al-Smadi et al. [20] used n-gram features and built SVM models for this subtask. Patra et al. [22] used WordNet information to find the frequency of each category in the hypernym tree of the grammatical object in the sentence and used its frequency for each category for building the CRF model.

The hand-crafted features like BOW and TF-IDF do not capture the word order information of a sentence. So, Mikolov et al. [15] proposed a mechanism of deriving vector representation of words using deep recurrent neural networks applied over a huge corpus. These word representations are also called word vectors. They not only capture the syntactic properties of the language, but also the semantic properties of the language. Blinov and Kotelnikov[7] used the skip-gram model to get word representations. The average of word vectors of words in a review sentence was used to represent a sentence. Further, the distance metric was used for identifying the category of a sentence. The Word2Vec based method was also used by Bilgin and Köktaş[26] for deriving document vectors and performing sentiment classification. Such word representations are very useful for smaller annotated datasets.

For ACD in Task 5 of SemEval-2016, Toh and Su[27] and Ruder et al. [8] also used Convolutional Neural Network (CNN) based architecture for this task. Toh and Su[27] utilized CNN-based features as input to the FFNN for unconstrained submission of English reviews. Khalil and EI-Beltagy[28] proposed an ensemble of CNN and SVM (with BOW features) for the '*restaurants*' domain. Xenos et al. [9] also used multiple ensemble classifiers (SVM) for each of the appearing- E and A tuples. Only the classifiers for which the confidence scores exceeded a particular threshold, determined the aspect category. Tamchyna and Veselovska[29] employed a simple LSTM based architecture, for this subtask from reviews of different languages and domains. However, their system did not show promising results as compared to baseline results.

Specifically, for Hindi ABSA Dataset, Akhtar et al. [17] had experimented with three classifiers, Decision tree, Naïve-Bayes, and SVM with lexicon features like -grams (basic, character, and non-contiguous). Among the results obtained for four major domains, Naïve Bayes and Decision Tree both showed the best results for two domains each.

In this paper, we target ACD subtask for reviews from Hindi ABSA Dataset released by Akhtar et al. [17]. Hindi is a rich morphological language and has a scarcity of resources for feature extraction. So, we derived features using word-embeddings and built the ensemble of 3 classifiers. We also feed these features to the Feed Forward Neural Network and compare the results.

2.2.Aspect Category Polarity

Determining the polarity of the review sentence is a simple classification problem. Here, we are more concerned about determining the polarity of each aspect category referenced in the review sentence. The set of categories are predefined for a particular domain but they may vary across domains. However, as aspect categories are not explicitly mentioned in the review sentence, such a determination relies upon two tasks- a) Finding the associated words which are closest to the aspect category b) Considering those words and their dependencies for building the classification model. Thus, we can determine the sentiment stated by the reviewer for each aspect category.

For sentiment analysis, there are a wide variety of features extracted and many different models built. In the last decade, the lexicon-based features were derived and used for classifying sentiments [30], [31], [32]. Wilson et al. [30] focused on assigning polarity to small phrases within reviews by selecting a lexicon and expanding it using a dictionary and thesaurus. Thet et al. [31] used SentiWordNet to derive prior sentiment scores using grammatical clause structure. Using a similar approach for micro-phrases, Musto et al. [33] used 4 different lexicons and experimented with different ways of getting sentiment scores of the tweets. Apart from this, Blitzer et al. [34] put forth a structural correspondence learning algorithm for domain adaptation by selecting the pivots features to link the two domains. The lexicon-based features [32] were used to build models using supervised learning for sentiment analysis.

At the basic level, the sentiment classification task was addressed by Hu and Liu [35] by identifying the orientation of opinion words in the review sentence. Ganu et al. [3] were the first to make use of supervised SVM-based models for sentiment classification at the sentencelevel. Further, as a part of Task 4 of SemEval2014 [2], mostly all submissions made use of the supervised machine learning approach for ACP subtask. To the best of our knowledge, Blinov and Kotelnikov[7] were the only onesto target this task using word vectors. This word representation was used to derive the average vector which could represent a sentence and ultimately represent each category-polarity combination, with a point in vector space. Further, with a simple distance measure, the polarity was assigned to a test sentence for a particular aspect category.

Further, for sentiment classification in ABSA, the machine learning models were preferred by several researchers. Castellucci et al. [4] made use of multi-kernel SVM in a one-versus-all setting. The different kernel functions used for ACP subtask include Partial Tree Kernel formulation and Polynomial kernel function. Kiritchenko et al. [4] also made use of multi-class SVM with one classifier for each aspect category. Their results reveal that the lexicon-based features had the highest contribution in improving the performance. Other than SVM, the Maximum entropy classifiers, Naïve Bayes, and many other classifiers are used for this purpose.

Recently, for simple document-based sentiment analysis, Kumar and Zymbler [2] used Glove word embeddings and n-gram features to classify the tweets into two classes- *positive* and *negative*. Among the SVM model, different ANN-based models, and the CNN model, the CNN models provided the best performance.

The Ensemble-based models proposed for polarity determination by Brun et al. [21] ranked first and reported accuracy of approximately 88% and 79% for English and French reviews respectively. The system employs methods to derive an aspect-centric representation of features and assigns polarity with the highest probability to a term or sentence using Conditional Random Field (CRF). Finally, the aspect category detected is also used as the feature with syntactic features from the parser. Xenos et al. [9] also used Ensemble of two SVM models and

reported more than 76% accuracy for both domains. The first model was trained using hand-crafted and lexicon-based features. The second model was built by deriving features from wordembedding.

The ABSA task was addressed for the dataset in different languages as well. For the Czech language, Hercig et al. [10] built a Maximum Entropy classifier with n-gram features and cluster-based features with tf-idf score, considering the whole sentence as context. For Hindi, the ACP subtask is addressed by Akhtar et al. [17] using three classifiers with a combination of lexical and polarity-based features. For polarity determination, the semantic orientation was defined for each token and used as features. Among the three classifiers, the Decision Tree and SVM-based approach gave better results than the Naïve Bayes classifier approaches.

For Arabic reviews, Al-smadi et al. [36] used SVM with frequent unigram and category-specific features to build the models for a similar task. In the same context, Guellil et al. [12] used Word2Vec based features with classical ML algorithms and Word2Vec and FastText based features with deep learning algorithms. Thus, they put forth the comparison of machine learning and deep CNN and LSTM based models for this task. The deep learning-based approach was also employed by Ruder et al. [8] for determining sentiment polarity for seven different languages. The basis for this approach was the use of a Convolutional Neural Network with aspect vector and sentence vector (derived from word embeddings) fed as input.

Thus, for ACP subtask, we prefer using a supervised machine learning approach. The features extracted for the classification task are based on pre-trained word vectors and a list based on domain-specific features extracted for each polarity level. The models used for experimentation include both the Ensemble of machine learning models and FFNN based models.

3. Methodology

The goal of the ACD subtask is to identify the presence of all aspect categories referenced into it. Thus, it is a multi-label classification problem. In this section, we describe the different ways of representing a sentence vector and the two models proposed by us using a sentence vector as features. We also extend this feature set for ACP subtask and experiment with the same two models. This problem is however a multi-class classification problem.

3.1.Sentence Vector

We retrieve the 300-dimensional word vectors of Hindi words from FastText. They are pre-trained vectors obtained by training on a huge corpus using Continuous Bag of Words (CBOW), with position weights. Thus, for each word, w_i the vector, $v_i$of 300-dimensions is retrieved from FastText. Here, v_i^j denotes the j^{th}component of it. We convert the review sentence, s (of length, l) using the word vector of dimension, d into sentence vector, SV_Sum using Equation (1). Equations (2), (3), and (4) are other representations of the review sentence used for experimentation. Here, the symbols $\| \; and \; \oplus$ represent the concatenation of vectors. The SV_Con represents the review sentence representation obtained by concatenating the first three sentence vector representations.

$$SV\_Sum(s) = \Big\|_{i=1}^{d} \sum_{j=1}^{l} v_i^j \tag{1}$$

$$SV\_Mul(s) = \Big\|_{i=1}^{d} \prod_{j=1}^{l} v_i^j \tag{2}$$

$$SV\_Min(s) = \Big\|_{i=1}^{d} \min_{i=1 \; to \; l} v_i^j \tag{3}$$

$$SV\_Con(s) = SV\_Sum \oplus SV\_Mul \; \oplus SV\_Min \tag{4}$$

3.2.Proposed ACD Models

Subtask 3 of ABSA is a multi-label classification task. Usually,for such multi-label classification, a separate classifier is used for each category (of that domain) to determine the aspect categories for a given sentence. This approach is called the binary relevance approach [17]. Thus, the sentence, *s*, is assigned category, *c* as given in Equation (5).

$$s_c = \begin{cases} 1, & if p(c|s) \geq t_1 \\ 0, & otherwise \end{cases} \quad (5)$$

If $|C| = 5$, a set of 5 binary classifier models need to be built, one for each category. Here, t_1 denotes the threshold for all reviews of a particular domain and s_c denotes the assignment of category, c to the sentence, s when s_c is set to 1. The value of this threshold t_1 is considered to be 0.5 for each classifier and each class when we make use *OneVersusRest* Classifier. Finally, we concatenate all aspect categories, c for which, s_c is set to 1 to obtain aspect categories for that review sentence, s. The two types of models proposed by us are the Ensemble-based model and Feed Forward Neural Network (FFNN) based model.

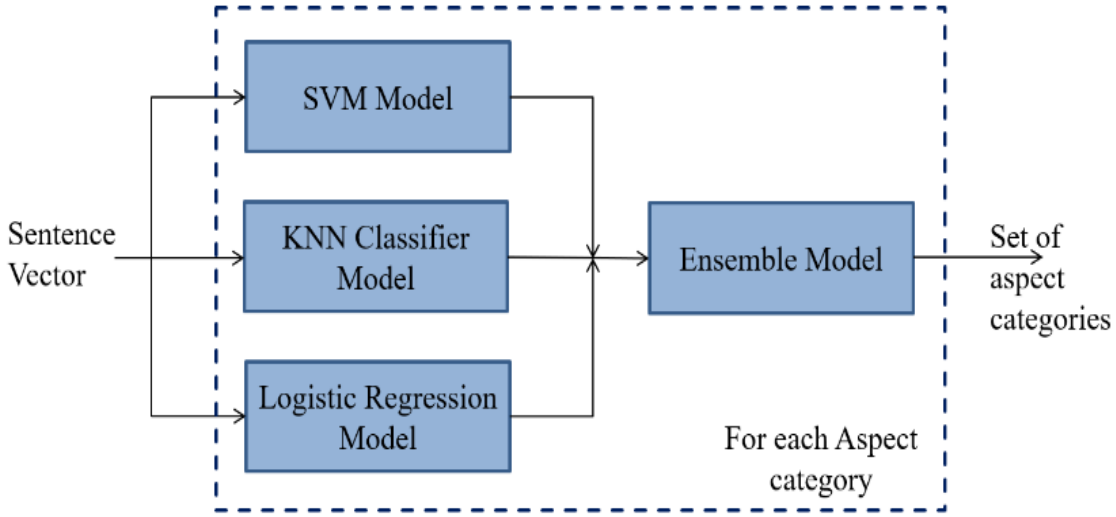
For a sample review, s from the test set, we retrieve the probability $p(c|s)$ for aspect category, c by using the binary classifier model. Thus, we get the probability of occurrence of all aspect categories for that domain. The reason for choosing the Ensemble model was that the performance of individual models was not appealing consistently across all the domains. To increase the performance, we choose the three base classifiers- SVM Classifier, KNN Classifier, and Logistic Regression based classifier to form an ensemble. The ensemble model is a combination of both linear and non-linear models. The block diagram of our ensemble-based proposed model is as depicted in Figure 3.

Further, we consider the linear combination of individual probability values from three classifiers for the same aspect category. Finally, we assign a category c, if this linear sum value exceeds the threshold, t_2. With the ensemble, the probability values of three models are used to assign the category c, to a sentence s using Equation (6).

$$s_c = \begin{cases} 1, if\ p_1(c|s) + p_2(c|s) + p_3(c|s) \geq t_2 \\ 0, \quad otherwise \end{cases} \quad (6)$$

The suffix, m in $p_m(c|s)$ denotes the model, m built for a category$c$, given a sentence s. Here, m=1 denotes the SVM Classifier model, m=2 denotes the KNN Classifier model and m=3 denotes Logistic Regression based classifier model.

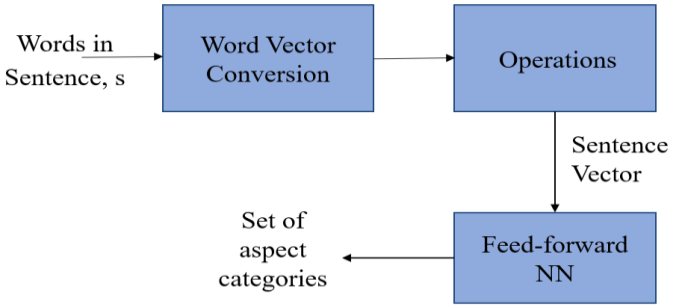
Figure 3. Block diagram of aspect category detection using the Ensemble model



The usage of the FFNN model for this ABSA subtask is as demonstrated in Figure 4. The sentence vector derived from words in the review sentence is fed as input to the model. The FFNN consists of two hidden layers, with their nodes activated using Rectified Linear Unit (ReLU) activation function. If there are n possible aspect categories, for the reviews from a particular domain, we build a model with n nodes in the output layer. The '*sigmoid*' activation function is used for nodes in the output layer to predict the probability of a particular aspect category.

For both the models, we derive the sentence vector for each review using Word-Vector conversion for words in the sentence and perform different operations onto them as explained above in section 3.1.

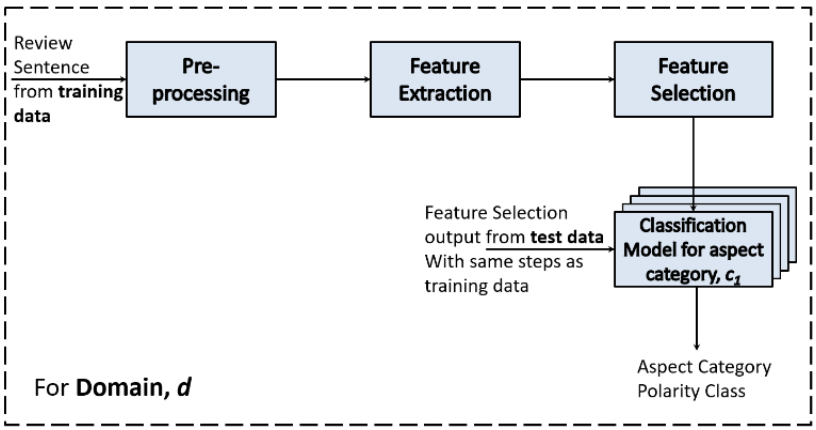
Figure 4.Block diagram of the Aspect category detection subtask using Feed Forward Neural Network



3.3.Feature Extension for ACP

Subtask 4 of ABSA is concerned with determining the polarity of each aspect category referenced in a review sentence. There are four polarity levels possible for aspect categories are- *positive*, *neutral*, *negative*, and *conflict*, as described above. Thus, given the review sentence with the aspect category, the goal is to predict the polarity associated with the aspect category. As each review sentence can have multiple aspect categories, we are supposed to find polarity for each aspect category. But each aspect category can have only one of the four polarity levels. So, it is modeled as a multi-class classification problem. However, to address this subtask, we build separate models for each category polarity for each of the four major domains. The diagrammatic representation of addressing ACP subtask with training and test data from each domain is represented in Figure 5.

Figure 5. Building Multi-class Classification models for each Aspect Category Polarity using training and test reviews from the corresponding domain, *d*



The sentence vector derived from a review sentence using word embeddings is further extended for ACP subtask using the following features.

Aspect Category based Features The reviewer may be having difference in polarity towards each aspect category referenced in the review sentence. To account for it, we deduce four words having maximum association with each aspect category and use the presence of these words in the review sentence as features.

N-gram Features The presence of all word unigrams and bigrams occurring in both training and test review sentences are used as features. Here, unigrams are tokens appearing individually in a review sentence. The bigrams are all pairs of successive tokens in it i.e all pairs of the form (w_k, w_{k+1}) from each review sentence. The actual number of features contributing to building the model from this set depends on the total number of reviews for a particular domain. Henceforth, we denote the unigrams features and bigrams features as Uni and Bi respectively.

Term Frequency-Inverse Document Frequency Features For the word unigrams and bigrams the term frequency and inverse document frequency score is also used as features. To account for this score in any case, we derive this score for both unigram and bigram features whenever such features are used to build the model.

We retrieve all the features described above to represent a set of features for each review sentence. We split the dataset for each of the four major domains using 3-fold cross-validation such that every time 2 folds are used for training and the remaining fold is used for testing the performance of the system. Then we proceed for building the model without feature selection.

3.4.Proposed ACP Models

We work with Ensemble Model and FFNN based Models for this ACP subtask. The proposed Ensemble-based model is a combination of the same three classifier models as the one used for the ACD subtask. It is a combination of SVM, KNN, and LoRclassifier models. The base classifiers are known to outperform for different text categorization tasks including sentiment classification [37]. As LoR and SVM are binary classifiers, we use them in a *One-versus-Rest* setting. As a result, one classifier is built for each polarity class by fitting it against all other classes.

To test the performance of the models built, we create an ensemble of the above three models using the following two methods-

- **Using summed Probability** Retrieving the probability values for each polarity label from three classifier models and summing them. Finally, the polarity label for which we have the maximum sum is assigned as the class label. Depending on the probability values for each polarity class, the class label is assigned binary value, as given in Equation (7). CP denotes the set of polarity labels, and $pr_1(c_i|s)$denotes the probability of assigning polarity label c_i to sentence s, using the first model in the ensemble. If the value assigned to s_x is 1, it denotes the sentence, *s* is assigned category polarity label, *x*, as it has the maximum value of the summed probability among all polarity labels. The results obtained with summed probability are indicated as E_SP.

$$s_x = \begin{cases} 1, & if \ \max\limits_{c_i \in CP}[pr_1(c_i|s) + pr_2(c_i|s) + pr_3(c_i|s)] \\ 0, & otherwise \end{cases} \quad (7)$$

- **Using majority voting**Here, we predict the category labels using each classifier model in the ensemble and assign the class label using majority voting. In case there is no clear majority, we use the summed probability method. As an alternative, we may compare the probability of each polarity label and assign the polarity with maximum value.The results obtained using majority voting are indicated as E_MV.

The Neural Network-based models are suited for complex classification problems having a large number of features. So, we work with features described in section 3.3 and build the model using single Feed Forward Neural Networks for each domain. The input layer consists of features extracted from a review sentence, with an aspect category for which the polarity label is under consideration. We make use of two hidden layers with the '*tanh*' activation function. The output layer consists of four neurons with '*softmax*' activation as there can be always any one of the four category polarity labels assigned to a sentence.

4. Results and Discussion

In this section, we brief upon the Hindi Aspect Category Dataset [17], the evaluation parameters used, the results obtained by our models for the ACD and ACP subtasks for the Hindi ABSA dataset.

4.1.Dataset

The ABSA Dataset released by Akhtar et al. [17] contains 5417 Hindi annotated review sentences from 12 different domains in XML format. Among these 12 domains, the 9 domains are grouped and named as *Electronics* domain [17]. Thus, the results of the aspect category dataset are demonstrated using four major domains- *Electronics*, *Mobile-apps*, *Travels*, and *Movies*. The aspect categories appearing for each domain are constrained to a specific set of features.

Each review sentence is annotated with the set of aspect categories referenced in it along with its sentiment polarity. The sample annotated review sentences from this dataset are as shown in Figure 6. It shows two review sentences from the *Mobile-apps* domain, with the second sentence having multiple aspect categories with opposite polarities assigned to them.

The distribution of aspect categories for reviews from each of the four major domains is as represented in Table 1. It can be observed that among the four domains, the Electronics domain contains more than 68.5% of the

aspect categories referenced. The number of categories into which the reviews in different domains are categorized vary from 4 to 6 across these domains.

Figure 6.Snapshot of Review from Hindi Aspect Category Dataset

```
<sentence polarity="pos" id="app_69">
    <text>और इसका सबसे अच्छा भाग यह है कि मूवीज की पूरी लाइब्रेरी को बिना किसी
        अतिरिक्त चार्ज के स्ट्रीम किया जा सकता है।</text>
  - <aspectCategories>
        <aspectCategory polarity="pos" category="misc"/>
    </aspectCategories>
</sentence>
<sentence polarity="pos" id="app_70">
    <text>60 रूपये प्रति माह में लाइव टीवी स्ट्रीमिंग और ऑन-डिमांड सेक्शन में उपलब्ध एक
        मैसिव मूवी लाइब्रेरी, यह बुरा सौदा नहीं है।</text>
  - <aspectCategories>
        <aspectCategory polarity="neu" category="price"/>
        <aspectCategory polarity="pos" category="misc"/>
    </aspectCategories>
</sentence>
```

The distribution of aspect categories for reviews from each of the four major domains is as represented in Table 1. It can be observed that among the four domains, the Electronics domain contains more than 68.5% of the aspect categories referenced. The number of categories into which the reviews in different domains are categorized vary from 4 to 6 across these domains.

Table. 1. Analysis of Hindi Aspect Category Dataset for four major domains based on the number of reviews having aspect category, *c* in it.

Aspect Category, *c*	**Number of Reviews having aspect category, *c* in**			
	Electronics Domain	**Mobile_apps Domain**	**Travels Domain**	**Movies Domain**
Design	524			
Ease of use	122	29		
Hardware	1797			
Misc	573	134	105	573
Price	228	11		
Software	370			
Gui		27		
Place			303	
Reachability			35	
Scenery			127	
Music				38
Performance				244
Story				35
Total	**3614**	**201**	**570**	**890**

Among the total 4930 review sentences in the Hindi Aspect Category Dataset having at least one aspect category, there is mention of 5275 aspect categories with their polarity. The distribution of the number of review sentences having either 1, 2, or 3 aspect categories across different domains is represented in Figure 7.We also analyze the whole dataset without considering the domain of the review sentence resulting in 13 aspect categories in all. We present the results of ACD and ACP with 3-fold cross-validation to enable comparison with the results from Akhtar et al. [17]. In the case of ACD, we retrieve stratified folds by considering the class labels as in *Aspect Category*. However, we retrieve stratified folds by considering the class labels as the combination of *Aspect Category* and *Polarity*.

Subtask 4 is concerning polarity determination of each referenced aspect category in a review sentence. So, the review sentences having no aspect categories are discarded. The four polarity levels for aspect categories are *positive*, *negative*, *neutral*, and *conflict*. A brief overview of the dataset concerningthe polarity distribution of aspect categories across domains is given in Figure 8.

From Figure 8, it can be observed that there exist around 45% to 62% of aspect categories with *positive* polarity in all domains except the *Movies* domain. Also, more than 85% of aspect categories from the whole dataset have *positive* or *neutral* polarity associated with them.

4.2.Feature Extraction

The quality of features extracted for building the machine learning model plays a vital role in the performance of the model. We are working with reviews in Hindi for the ACD subtask. Hindi being a resource-

Figure 7. Distribution of the number of aspect categories per review sentence in Hindi ABSA Aspect Category Dataset

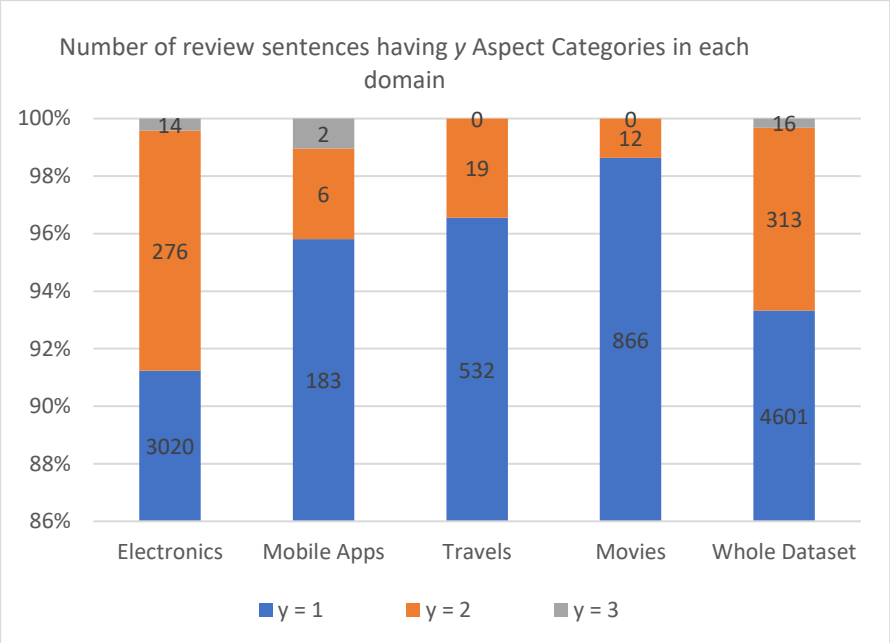
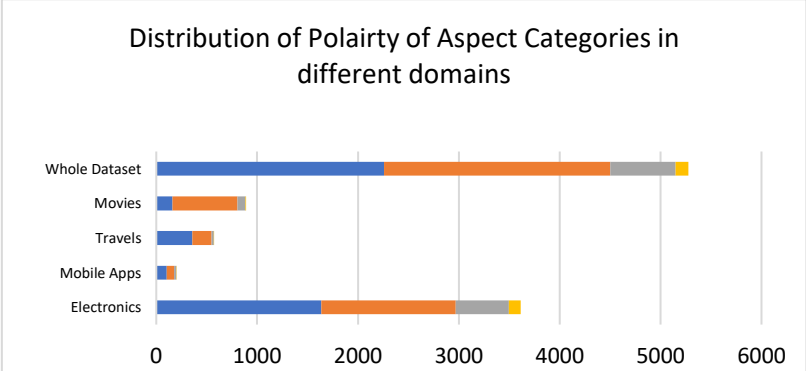


Figure 8. Polarity Distribution for each of the four domains in Aspect Category Dataset of Hindi ABSA Task



	Electronics	Mobile Apps	Travels	Movies	Whole Dataset
Positive	1635	104	358	160	2257
Neutral	1335	74	190	645	2244
Negative	524	22	21	79	646
Conflict	120	1	1	6	128

scarce language, we saw that the lexical features are most helpful. In addition to this, we also derive the word embedding-based features for improved performance. We use the pre-trained word vectors (trained with large Hindi corpus) provided by FastText as feature representations for the words. We employ these word vectors to derive a sentence vector, for each review sentence in the dataset as explained in the section 3.1.

4.2.1.Aspect Category Detection

The word embedding of each word, wi is a vector vi of 300-dimensions and it is obtained from FastText. The sentence vector is derived using Equations (1), (2), (3), and (4). The sentence vector obtained using Equations 1, 2, and 3 is 300D, and using Equation (4), it is 900D.

4.2.2.Aspect Category Polarity

The features used for ACP models include a sentence vector derived using FastText for representing a review sentence. For the Aspect Category Polarity subtask, we choose sentence vector derived using Equation (1). They are further augmented by finding four words and their features having maximum association with each aspect category. We also derive N-gram and Term frequency and inverse-document features from the review sentence for the ACP subtask.

4.2.3.Feature Selection for ACP

For ACP subtask, we prefer employing feature selection methods for domains having a large number of unigram and bigrams. We make use of the ANOVA F-test to filter out the selected number of features. As the feature values for all features extracted are numerical, we make use of the ANOVA F-test for feature selection. The number of selected features for ANOVA for a particular domain depends upon the actual count of original features, F retrieved for that domain. If F is large, a smaller percentage of F is used as a count of selected features, whereas, if F is small, a larger percentage of F is used as a count of selected features. We retrieve the results of ACP, both with and without feature selection.

4.3.Performance Evaluation

The performance measure mostly used for any classification task is '*accuracy*'. However, being a multi-label classification task, the '*F-score*' evaluation measure is used. The F-score is a micro averaged F-score. If TP_x denotes the number of True Positives for the aspect category, *x*, PCP_x denotes Predicted Condition Positive for the aspect category, *x* and ACP_x denotes Actual Condition Positive for the aspect category, *x*, wecompute the micro-Precision, Recall and F-score by using Equation (8), (9) and (10).

$$Precision, P_{micro} = \frac{\sum_{x \in C} TP_x}{\sum_{x \in C} PCP_x} \quad (8)$$

$$Recall, R_{micro} = \frac{\sum_{x \in C} TP_x}{\sum_{x \in C} ACP_x} \quad (9)$$

$$F - score, F_{micro} = \frac{2 * P_{micro} * R_{micro}}{(P_{micro} + R_{micro})} \quad (10)$$

The goal of the ACP Subtask of ABSA is to find the polarity of aspect categories referenced in the review sentence. The Hindi ABSA dataset contains four polarity classes- *positive*, *negative*, *neutral,*and *conflict*. The performance measure mostly used for a multi-class classification task is '*Accuracy*'. It is defined as the ratio of the total number of correctly classified samples to the total number of samples considered or tested. Thus, It would correspond to the ratio of the sum of diagonal terms in the confusion matrix of multi-class classification problem over the total size of the test sample. If *CP* is the set of possible category polarities, TP_x denotes the number of true positives for some polarity *x*, where $x \in CP$, Accuracy is defined as given in Equation (11).

$$Accuracy, A == \frac{\sum_{x \epsilon CP} TP_x}{Total\ size\ of\ the\ test\ sample} \quad (11)$$

4.4.Experimental Results

In this sub-section, we put forth the results attained, for detecting aspect categories for Hindi Aspect Category Dataset [17] using 3-fold cross-validation. For experimentation, we make use of Python as a programming language with Python libraries, Sklearn, Tensorflow, Keras, etc. We compare the state-of-the-art results with our proposed Ensemble and FFNN models.

4.4.1.Ensemble Model for ACD

To address the ACD subtask, instead of using a single classifier model, we use the ensemble of three classifier models- SVM, KNN, and the Logistic Regression model. Though the Ensemble model involves additional

overhead, we propose this model to derive a more generic model which yields relatively better results across all the domains.

The review sentence is represented using different variations of sentence vectors as mentioned in the section 3.1. The sentence vector itself is used as a feature vector to build the classification model. The F-score obtained for test reviews of the *Electronics* domain, using four types of sentence vectors are depicted in Table 2.

Table 2 Aspect Category Detection F-score results of test dataset from *Electronics* domain with three individual classifiers and the ensemble with different sentence vector representations.

Models Domain *Electronics* **Sentence Vector**	**SVM Model**	**KNN Classifier Model**	**Logistic Regression Model**	**Ensemble Model**
SV_Sum	50.31	49.73	65.13	**66.83**
SV_Min	51.91	49.59	53.7	56.04
SV_Mul	51.91	34.13	51.92	46.35
SV_Con	49.47	50.15	65.32	**66.9**

From Table 2, it can be observed that among all the four models, the Ensemble Model with SV_Con and SV_Sum as sentence vector provide mostly similar and maximum F-score. Among the three basic classification models, the Logistic Regression model shows maximum performance for the *Electronics* domain. However, the performance is still somewhat less than the Ensemble model in a similar setting. However, the F-score for the *Mobile_apps* domain had Support Vector Machine giving better results among the three basic classification models considered. The average F-score results obtained for the *Mobile_apps* domain with 3-fold cross-validation are as tabulated in Table 3.

Table 3 Aspect Category Detection F-score results of test dataset from *Mobile_Apps* domain with three individual classifiers and the ensemble with different sentence vector representations.

Models Domain *Mobile_Apps* **Sentence Vector**	**SVM Model**	**KNN Classifier Model**	**Logistic Regression Model**	**Ensemble Model**
SV_Sum	68.15	61.28	63.71	**69**
SV_Min	66.14	60.72	67.51	67.33
SV_Mul	65.1	63.55	67.4	67.64
SV_Con	67.5	61.82	64.91	**68.14**

From Table 2 and 3, it is observed that the Ensemble model with the SV_Sum as sentence vector provides a rise in F-score of a maximum of 1.67% over that of Ensemble models with other sentence vector representation methods. In the case of both *Travels* and *Movies* domain also, the Ensemble model with SV_Sum gives the best results as compared to the other three review sentence representation methods. Overall, it was observed that none of the classifiers performed well across all the domains. Thus, the linear combination of these models is our proposed model for this classification task. The Final F-score results obtained using different models for each domain with SV_Sum as sentence vector are as tabulated in Table 4.

Table 4 F-score of ACD Subtask obtained using different models and their *Ensemble* for four domains

Models **Domain**	**SVM Model**	**KNN Classifier Model**	**Logistic Regression Model**	**Ensemble Model**
Electronics	50.31	49.73	65.13	66.83
Mobile Apps	68.15	61.28	63.71	69

Travels	53.09	56	59.77	63.75
Movies	64.74	61.14	72.77	73.49

The complete set of results are obtained using 3-fold cross-validation. The value of threshold t_2 was empirically set to 1.0 in the case of the Ensemble model for maximum performance.

From Table 4, it is observed that the F-score of the Ensemble model is superior to the three basic models when models are built individually for each domain. However, if we build a single model using the whole dataset, the obtained F-score is 63.36% which is the minimum among the F-score of individual domains. The F-score decreases and the time required to build the model is also more than 10 times required for building models for an individual domain. One of the reasons for this is that when the whole dataset is considered,a total of 13 models are built (one model for each aspect category).

4.4.2.FFNN Model for ACD

The other model proposed by us for the ACD subtask of ABSA is based on the Feed Forward Neural Network. It consists of input, which is a sentence vector representation SV_Sum, and multiple outputs, each of which represents the presence or absence of a particular aspect category, for the review sentence of that particular domain. The model was built using two hidden layers with a '*ReLU*' activation function. The output of nodes in the output layer corresponds to the probability that the review sentence references a particular aspect category. We use the '*sigmoid*' activation function for output layer nodes. The hyper-parameters of the neural network are empirically set and are as tabulated in Table 5.

Table 5 Optimal hyper-parameters for FFNN model for Aspect Category Detection

Parameter Name	**Value**
No. of Hidden Layers	02
No. of neurons in the first hidden layer	200
No. of neurons in the second hidden layer	100
Initial Learning rate	0.003
Batch size	30
No. of training epochs	150
Regularization parameter	0.002
Optimizer	Adam

The learning rate is decreased using the exponential decay function, as the learning progresses. The L2-regularization parameter is set to 0.002 to reduce the loss. We obtain the domain-wise F-score measure, which is as tabulated in Table 6. The maximum F-score of 72.5% is obtained for the *Movies* domain. However, for reviews from the *Travels* domain, this F-score of 56.7% using FFNN models is much less than that obtained by the Ensemble model (63.75%).

Table 6 F-score of ACD Subtask obtained using FFNN for four domains.

Domain	**Precision**	**Recall**	**F-score**
Electronics	68.1	65.9	67
Mobile Apps	70.1	68.5	69.3
Travels	59.5	55.6	56.7
Movies	76.7	68.8	72.5

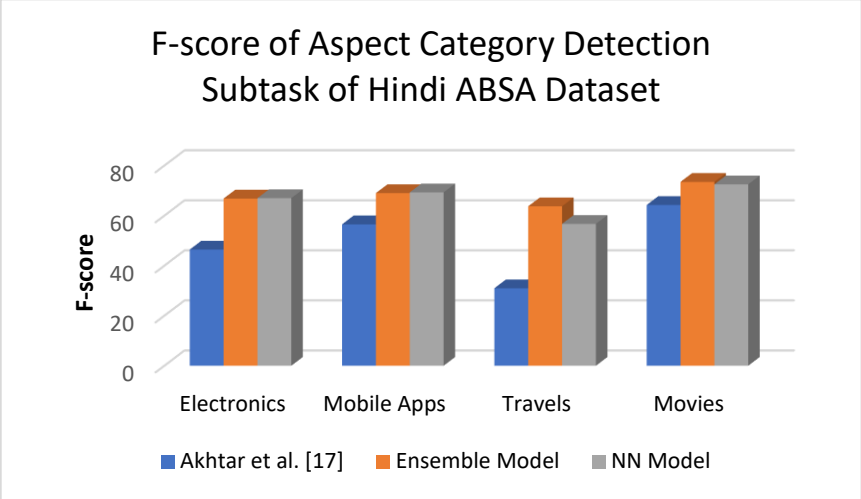
4.4.3.Comparison of ACD Results

Here, we compare the aspect category dataset results obtained by us using Ensemble Models and FFNN with the state-of-the-art results. The F-score comparison is as tabulated in Figure 9. The results obtained by us using 3-fold cross-validation show considerable improvement in the F-score for all the domains than the results shared by Akhtar et al. [17]. With our proposed Ensemble model, we attain a maximum F-score of 63.75% and 73.49% for the *Travels* and *Movies* domain respectively. For both *Electronics* and *Mobile_Apps* domain, both Ensemble and FFNN model result in nearly the same F-score values.

4.4.4.Ensemble Model for ACP

We propose the Ensemble of three classifier models- K-Nearest Neighbor (KNN), Logistic Regression classifier (LoR), and Support Vector Machine (SVM) models for this subtask. The two ensemble methods employed for getting prediction results over samples from the test set using output from three base classifiers are as explained in the section 3.4.

Figure 9. Comparison of our ACD F-score of proposed Models with state-of-the-art results



The comparison of results using these two methods with all the features and without feature selection is stated in Table 7. In this table, E_SP and E_MV denote the accuracy of the ensemble over test data using Summed probability and Majority voting method respectively. It can be observed that the ensemble results with summed probability values are consistently maximum or just less than 0.50 of the maximums for every domain. So, we choose the summed probability-based method for assigning polarity to a review sentence for a particular aspect category. In Table 7, the accuracy values which are within 0.5 of maximum are also shown in bold.

The above results in Table 7 are obtained using the complete set of features and obtained with 3-fold cross-validation. The average results obtained with a partial set of features using the Ensemble model for each of the four major domains are as tabulated in Table 8. So, after comparison of results, it can be observed that the resultsobtained when choosing all features (SV + Uni + Bi) together dominates and so are retained for comparison with the state-of-the-art results.

Table 7 Accuracy results of Ensemble over test data using Summed probability and Majority voting method with the complete set of features without feature selection.

Domain	**SV + Uni + Bi**				
	SVM	**KNN**	**LoR**	**E_SP**	**E_MV**
Electronics	61.46	61.57	**69.70**	**69.15**	67.85
Mobile Apps	56.34	55.42	59.08	**63.96**	**64.40**
Travels	68.54	67.31	**72.59**	**72.58**	71.01
Movies	77.09	75.96	**78.54**	**79.00**	**79.33**

Table 8 Accuracyresults obtained using different Machine Learning Models for Aspect Category Polarity Subtask of Hindi ABSA Dataset

Domain	**SV**				**SV + Uni**			
	SVM	**KNN**	**LoR**	**E_SP**	**SVM**	**KNN**	**LoR**	**E_SP**
Electronics	60.35	59.91	**65.85**	64.35	52.85	61.57	**69.7**	69.29
Mobile Apps	**52.05**	50.88	50.44	**52.42**	51.03	55.42	59.08	**62.01**

Travels	62.74	64.5	70.12	**70.3**	62.56	67.31	72.59	**72.76**
Movies	**72.47**	66.85	71.91	72.23	72.48	75.96	78.54	**78.77**

4.4.5.FFNN Model for ACP

In addition to the Ensemble-based approach, we also work with the Neural Network-based model.The performance of the system with a variety of features using a feed-forward neural network-based approach is as demonstrated in Table 9. From the table, it can be observed that when we do not employ feature selection, the combination of all features dominates and produces more or less maximum accuracy for all domains except the *Electronics* domain. Whereas, when we employ feature selection the sentence vector-based features dominate and produce near to maximum accuracy for all domains except *Electronics* and *Mobile-Apps*domain. It was only for the *Electronics* domain, that in both feature selection and without feature selection cases, the unigram features dominate. So, to make a comparison we choose the accuracy results with all features (SV + Uni + Bi) without any feature selection.

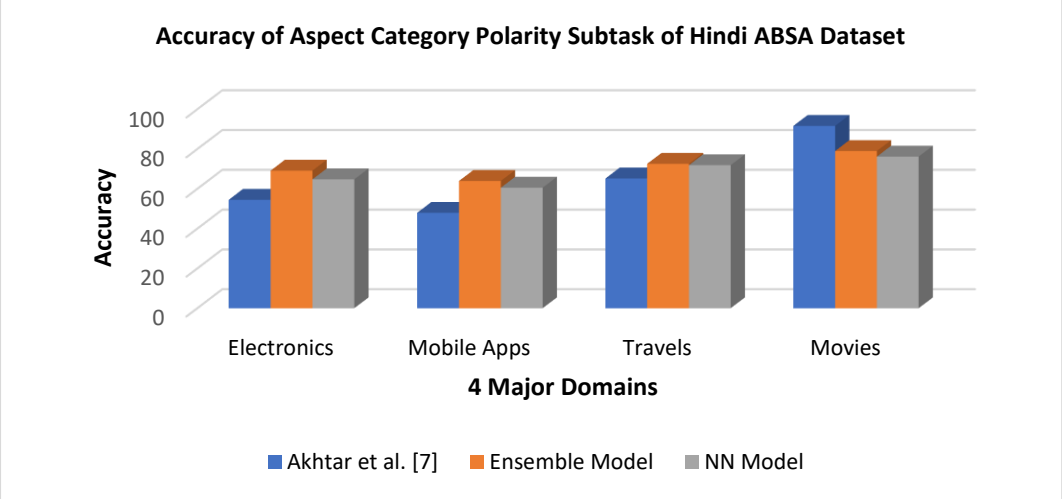
Table 9 Accuracyresults obtained using NN Models for Aspect Category Polarity Subtask of Hindi ABSA Dataset

Domain	**Without Feature Selection**				**With Feature Selection**			
	SV	**Uni**	**Uni + Bi**	**SV + Uni + Bi**	**SV**	**Uni**	**Uni + Bi**	**SV + Uni + Bi**
Electronics	63.37	**67.13**	66.93	64.78	64.08	**66.99**	66.91	64.67
Mobile Apps	58.57	54.98	59.49	**60.59**	56.08	**60.51**	55.54	57.00
Travels	71.18	67.66	69.07	**71.88**	**71.18**	67.66	69.59	70.65
Movies	75.96	75.97	**76.53**	76.18	**77.44**	75.96	76.74	76.74

4.4.6.Comparison of ACP Results

The comparison of our proposed Aspect category polarity results with the best state-of-the-art results is as represented in Figure 10. It can be observed that our Ensemble model provides an increase in accuracy in the range of 7% to 14% for all three domains except the *Movies* domain over best state-of-the-art results. The accuracy results of Neural Network-based models for all the domains is less than that obtained by the Ensemble-based Model.

Figure 10. Comparison of our ACP Accuracy score of proposed Models with state-of-the-art results



5. Conclusion and future work

In this paper, we propose Ensemble-based models for ACD and ACP subtask of Hindi Aspect Based Sentiment Analysis. As the subtask is for Hindi Reviews, we do not prefer extracting features for each review explicitly, but work with four ways of representing a review sentence using embedding vectors. Among the

different sentence representation methods, SV_Sum based representation provides superior or comparable performance across all the domains using Ensemble models than FFNN based models. Our proposed Ensemble-based Model provides an improvement of 9% to 32% F-score among 4 domains over state-of-the-art results.For ACP subtask as well, our Ensemble model provides an increase in accuracy in the range of 7% to 14% for all three domains except the *Movies* domain over best state-of-the-art results. For this ACP subtask, we propose using the combination of hand-crafted features with word embedding features for improved performance. Such analogy may be utilized and tested for ABSA subtasks in other languages.In the future, we also look forward to improving the performance by employing transfer learning-based techniques.

6. Acknowledgements

This work is not supported fully or partially by any funding organization or agency.

References

1. M. Hu and B. Liu, "Mining Opinion Features in Customer Reviews," in Proceedings of The 19th National Conference on Artificial Intelligence - (AAAI'04), 2004, pp. 755–760, [Online]. Available: http://dl.acm.org/citation.cfm?id=1597148.1597269.
2. M. Pontiki, D. Galanis, J. Pavlopoulos, H. Papageorgiou, I. Androutsopoulos, and S. Manandhar, "SemEval-2014 Task 4: Aspect Based Sentiment Analysis," in Proceedings ofthe 8th International Workshop on Semantic Evaluation, 2015, no. SemEval, pp. 27–35, doi: 10.3115/v1/s14-2004.
3. G. Ganu, N. Elhadad, and A. Marian, "Beyond the Stars : Improving Rating Predictions using Review Text Content," Twelfth Int. Work. Web Databases (WebDB 2009) J, vol. 9, no. WebDB, pp. 1–6, 2009, [Online]. Available: http://www.dbmi.columbia.edu/noemie/ursa.
4. G. Castellucci and S. Filice, "UNITOR : Aspect Based Sentiment Analysis with Structured Learning UNITOR : Aspect Based Sentiment Analysis with Structured Learning," in Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014), 2014, pp. 761–767, doi: 10.3115/v1/S14-2135.
5. T. Brychcín, M. Konkol, and J. Steinberger, "UWB: Machine Learning Aproach to Aspect-Based Sentiment Analysis," in Proceedings of the 8th international workshop on semantic evaluation (SemEval 2014), 2014, pp. 817–822.
6. S. de Kok, L. Punt, R. van den Puttelaar, K. Ranta, K. Schouten, and F. Frasincar, "Review-aggregated aspect-based sentiment analysis with ontology features," Prog. Artif. Intell., vol. 7, no. 4, pp. 295–306, 2018, doi: 10.1007/s13748-018-0163-7.
7. P. Blinov and E. Kotelnikov, "Blinov : Distributed Representations of Words for Aspect-Based Sentiment Analysis at SemEval 2014," in Proceedings of the 8th international workshop on semantic evaluation (SemEval 2014), 2014, pp. 140–144.
8. S. Ruder, P. Ghaffari, and J. G. Breslin, "INSIGHT-1 at SemEval-2016 Task 5: Deep learning for multilingual aspect-based sentiment analysis," SemEval 2016 - 10th Int. Work. Semant. Eval. Proc., pp. 330–336, 2016, doi: 10.18653/v1/s16-1053.
9. D. Xenos, P. Theodorakakos, J. Pavlopoulos, P. Malakasiotis, and I. Androutsopoulos, "AUEB-ABSA at SemEval-2016 task 5: Ensembles of classifiers and embeddings for Aspect Based Sentiment Analysis," SemEval 2016 - 10th Int. Work. Semant. Eval. Proc., pp. 312–317, 2016, doi: 10.18653/v1/s16-1050.
10. T. Hercig, T. Brychc´ın, L. Svoboda, M. Konkol, and J. Steinberger, "Unsupervised Methods to Improve Aspect-Based Sentiment Analysis in Czech," vol. 20, no. 3, pp. 365–375, 2016, doi: 10.13053/CyS-20-3-2469.
11. S. Kumar and M. Zymbler, "A machine learning approach to analyze customer satisfaction from airline tweets," J. Big Data, vol. 6, no. 1, pp. 1–16, 2019, doi: 10.1186/s40537-019-0224-1.
12. Guellil et al., "A Semi-supervised Approach for Sentiment Analysis of Arab(ic+izi) Messages: Application to the Algerian Dialect," SN Comput. Sci., vol. 2, no. 2, pp. 1–18, 2021, doi: 10.1007/s42979-021-00510-1.
13. S. Kiritchenko et al., "NRC-Canada-2014 : Detecting Aspects and Sentiment in Customer Reviews," no. SemEval, pp. 437–442, 2014.
14. S. Vicente, X. Saralegi, and R. Agerri, "Elixa: A modular and flexible ABSA platform," arXiv, 2017, doi: 10.18653/v1/s15-2127.
15. T. Mikolov, I. Sutskever, K. Chen, G. Corrado, and J. Dean, "Distributed Representations of Words and Phrases and their Compositionality," Adv. Neural Inf. Process. Syst., pp. 3111–3119, 2013.
16. S. Ahmad, M. Z. Asghar, F. M. Alotaibi, and I. Awan, "Detection and classification of social media-based extremist affiliations using sentiment analysis techniques," Human-centric Comput. Inf. Sci., vol. 9, no. 1, 2019, doi: 10.1186/s13673-019-0185-6.

17. S. Akhtar, A. Ekbal, and P. Bhattacharyya, "Aspect Based Sentiment Analysis : Category Detection and Sentiment Classification for Hindi," in International Conference on Intelligent Text Processing and Computational Linguistics, 2016, pp. 246–257.
18. M. Ghosh and G. Sanyal, "An ensemble approach to stabilize the features for multi-domain sentiment analysis using supervised machine learning," J. Big Data, vol. 5, no. 1, 2018, doi: 10.1186/s40537-018-0152-5.
19. O. Al-harbi, "Classifying Sentiment of Dialectal Arabic Reviews : A Semi-Supervised Approach," vol. 16, no. 6, 2019.
20. M. Al-smadi, O. Qawasmeh, M. Al-ayyoub, Y. Jararweh, and B. Gupta, "Deep Recurrent neural network vs. support vector machine for aspect-based sentiment analysis of Arabic hotels' reviews," J. Comput. Sci., 2017, doi: 10.1016/j.jocs.2017.11.006.
21. C. Brun, J. Perez, and C. Roux, "XRCE at SemEval-2016 task 5: Feedbacked ensemble modelling on syntactico-semantic knowledge for Aspect Based Sentiment Analysis," SemEval 2016 - 10th Int. Work. Semant. Eval. Proc., pp. 277–281, 2016, doi: 10.18653/v1/s16-1044.
22. B. G. Patra, S. Mandal, D. Das, and S. Bandyopadhyay, "JU _ CSE : A Conditional Random Field (CRF) Based Approach to Aspect Based Sentiment Analysis," pp. 370–374, 2014.
23. K. Schouten, O. Van Der Weijde, F. Frasincar, and R. Dekker, "Supervised and Unsupervised Aspect Category Detection for Sentiment Analysis With Co-Occurrence Data," pp. 1–13, 2017.
24. Kumar, S. Kohail, A. Kumar, A. Ekbal, and C. Biemann, "IIT-TUDA at SemEval-2016 task 5: Beyond Sentiment lexicon: Combining domain dependency and distributional semantics features for aspect based sentiment analysis," SemEval 2016 - 10th Int. Work. Semant. Eval. Proc., pp. 1129–1135, 2016, doi: 10.18653/v1/s16-1174.
25. T. Álvarez-López, J. Juncal-Martínez, M. Fernández-Gavilanes, E. Costa-Montenegro, and F. J. González-Castaño, "GTI at SemEval-2016 Task 5: SVM and CRF for aspect detection and unsupervised aspect-based sentiment analysis," in SemEval 2016 - 10th International Workshop on Semantic Evaluation, Proceedings, 2016, pp. 306–311, doi: 10.18653/v1/s16-1049.
26. M. Bilgin and H. Köktaş, "Sentiment analysis with term weighting and word vectors," Int. Arab J. Inf. Technol., vol. 16, no. 5, pp. 953–959, 2019.
27. Z. Toh and J. Su, "NLANGP: Supervised Machine Learning System for Aspect Category Classification and Opinion Target Extraction," vol. 14, no. SemEval, pp. 496–501, 2015, doi: 10.18653/v1/s15-2083.
28. T. Khalil and S. R. El-Beltagy, "NileTMRG at SemEval-2016 task 5: Deep convolutional neural networks for aspect category and sentiment extraction," SemEval 2016 - 10th Int. Work. Semant. Eval. Proc., pp. 271–276, 2016, doi: 10.18653/v1/s16-1043.
29. Tamchyna and K. Veselovská, "UFAL at SemEval-2016 task 5: Recurrent neural networks for sentence classification," SemEval 2016 - 10th Int. Work. Semant. Eval. Proc., pp. 367–371, 2016, doi: 10.18653/v1/s16-1059.
30. T. Wilson, J. Wiebe, and P. Hoffmann, "Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis," in Proceedings of the conference on human language technology and empirical methods in natural language processing, 2005, pp. 347–354.
31. T. T. Thet, J. Na, and C. S. G. Khoo, "Aspect-based sentiment analysis of movie reviews on discussion boards," J. Inf. Sci., no. November, 2010, doi: 10.1177/0165551510388123.
32. J. Rojratanavijit, P. Vichitthamaros, and S. Phongsuphap, "Acquiring sentiment from twitter using supervised learning and lexicon-based techniques," Walailak J. Sci. Technol., vol. 15, no. 1, pp. 63–80, 2018, doi: 10.48048/wjst.2018.2731.
33. C. Musto, G. Semeraro, and M. Polignano, "A comparison of lexicon-based approaches for sentiment analysis of microblog," in CEUR Workshop Proceedings, 2014, vol. 1314, pp. 59–68.
34. J. Blitzer, M. Dredze, and F. Pereira, "Biographies, bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification," in ACL 2007 - Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics, 2007, pp. 440–447.
35. M. Hu, B. Liu, and S. M. Street, "Mining and Summarizing Customer Reviews," 2004.
36. M. Al-Smadi, M. Al-Ayyoub, Y. Jararweh, and O. Qawasmeh, "Enhancing Aspect-Based Sentiment Analysis of Arabic Hotels' reviews using morphological, syntactic and semantic features," Inf. Process.
37. R. Xia, C. Zong, and S. Li, "Ensemble of feature sets and classification algorithms for sentiment classification," Inf. Sci. (Ny)., vol. 181, no. 6, pp. 1138–1152, 2011, doi: 10.1016/j.ins.2010.11.023