An Efficient Aspect based Sentiment Analysis Model by the Hybrid Fusion of Speech and Text Aspects

Maganti Syamala[1]
Research Scholar, Department of Computer Science and Engineering, Annamalai University, Annamalai Nagar, Chidambaram, Tamil Nadu 608002, India
Assistant Professor, Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation Vaddeswaram, AP, India

N.J.Nalini[2]
Department of Computer Science and Engineering
Annamalai University
Annamalai Nagar, Chidambaram
Tamil Nadu 608002
India

***Abstract*—Aspect-based Sentiment Analysis (ABSA) is treated to be a challenging task in the domain of speech, as it needs the fusion of acoustic features and Linguistic features for information retrieval and decision making. The existing studies in speech are limited to speech and emotion recognition. The main objective of this work is to combine acoustic features in speech with linguistic features in text for ABSA. A deep learning and language model is implemented for acoustic feature extraction in speech. Different variants of text feature extraction techniques are used for aspect extraction in text. Trained Lexicons, Latent Dirichlet Allocation (LDA) model, Rule based approach and Efficient Named Entity Recognition (E-NER) guided dependency parsing approach has been used for aspect extraction. Sentiment with respect to the extracted aspect is analyzed using Natural Language Processing (NLP) techniques. The experimental results of the proposed model proved the effectiveness of hybrid level fusion by yielding improved results of 5.7% WER and 3% CER when compared with the traditional baseline individual linguistic and acoustic feature models.**

***Keywords*—Acoustic; aspect-based sentiment analysis; decision making; emotion; extraction; hybrid; lexicon; linguistic; natural language processing; speech**

I. Introduction

Sentiment analysis or opinion mining is the area of study in NLP where it helps to analyze the polarity with respect to the given context. Sentiment analysis depicts the state-of-art-of-mind to automate the process of analyzing the opinion, emotion, polarity, appraisal, interest, ideology, attitude, feelings towards an entity. Sentiment analysis plays an important role in our daily lives for analysis and decision making. In most of the existing studies, sentiment analysis is been carried out on text and the performance is been differentiated by varying the type of linguistic features extracted from text. The features on text are generally called as linguistic features and play a very crucial role in sentiment analysis. Due to the tremendous growth of data in World Wide Web, now-a-days traditional and web-based surveys are been replaced by sentiment analysis [1]. As WWW is a combination of text, audio and video, there is a need for analysis of sentiment on multimodal data. Feature extraction for sentiment analysis will be differed for different types of input like text, audio and video. The field of sentiment analysis in NLP had gained its popularity by implementing on text. By the evolution of massive data, research is been expanded and now it's confined not only to text but also had gained its popularity in different modalities. When sentimental analysis came into picture, it's been carried out only on text using NLP and machine learning techniques, where the polarity of the given document or sentence is classified as either positive, negative or neutral [1]. Next era of sentiment analysis is aspect-based sentiment analysis (ABSA) and had gained its popularity in recommender systems. Most of the recommender systems that used ABSA have identified the sentiment with respect to the aspect in the given text. Parts-of-Speech (POS) tagging was one of the widely used aspect identification technique for ABSA [4]. In this paper, aspect-based sentiment analysis was been carried out by combining both audio and text features.

Most of the research so far carried out on audio data is confined to speech analysis and emotion recognition. In the existing studies [6], various acoustic features are analyzed and are classified for speech emotion recognition. Identifying sentiment in speech is a challenging task because of following reasons.

- Even though both the terms emotion and sentiment express feelings with respect to the context but the way they are analyzed is different.
- Emotion is the one that can be analyzed in speech by means of various acoustic features and prosodic features like pitch, intensity, energy, loudness etc. Whereas in text the sentiment is defined as an adjective that qualifies the respective noun.
- There is a difficulty to map emotion in speech with parts-of-speech in text for analyzing the sentiment. Even though there are many existing studies carried out on speech for sentiment analysis, the work is limited in analyzing only the emotion in speech like happy, sad, angry, fear and etc.; but not the positivity and negativity in the given context.

As speech and text features are different, so there is a need to bridge the gap between them to perform sentiment analysis. Speech in call-centers and text in recommender systems has gained its popularity in the field of sentimental analysis [15]. Fig. 1 depicts the sentiment analysis model by considering bi-modal speech and text features.

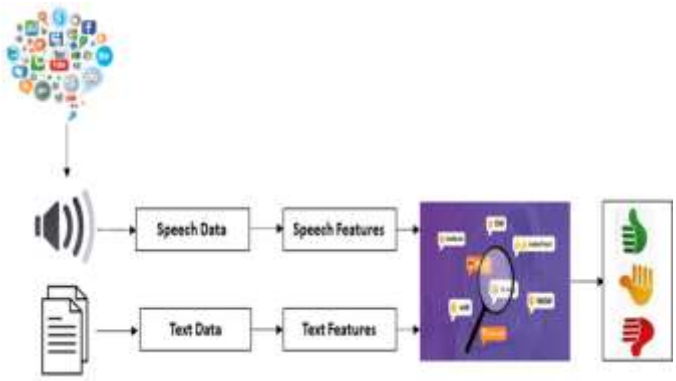

Fig. 1. Speech and Text based Bi-modal Sentiment Analysis.

The main contributions of the proposed work are:

- The importance of linguistic and acoustic features for ABSA is analyzed.
- A hybrid level fusion of acoustic and linguistic features for ABSA is evaluated using Word Error Rate (WER) metric and machine learning algorithms.
- The obtained results from proposed combined model are validated with the individual implementations of speech and text-based sentiment analysis.

II. Motivation

The field of sentiment analysis is catching everyone's attention in marketing, corporate and academia by executing the tasks in an easy and efficient manner. But most of the traditional frameworks are confined to work only either on text or audio or video. There is very limited study carried out on multimodal data. Now-a-days, sentiment is been analyzed as aspect-based sentiment analysis and major limitations are been identified in feature extraction and sentiment related aspect category identification. So, this made my work to drive towards implementing aspect-based sentiment analysis on multimodal speech and text data. Identification of sentiment with respect to the aspect helps to improve quality of service when compared with document and sentence level sentiment classification.

Now-a-days, tremendous growth of data available in social media and online commercial websites made everyone to provide online reviews demonstrated as a video in YouTube. Previously, consumers used take their decision for any purchase by analyzing the text reviews given by the customers [16]. In some cases, like where there is no customer who had already bought the product, there will be no rating and review provided for that product. In such cases consumer is not in a state to make a decision whether to go for it or not. So, this made me to develop an aspect-based sentiment analysis model on YouTube review data for improving the quality of service to consumers.

III. Related Work

The main objective of the proposed model is to analyze Aspect based sentiment analysis by combining both linguistic and acoustic features. Acoustic feature extraction techniques and Linguistic feature extraction techniques are applied for feature extraction [14] on YouTube product review dataset. The base line models are implemented by considering individual linguistic and acoustic features, validated using machine learning algorithms. A hybrid level fusion of acoustic and linguistic features for ABSA yields improved results when measured in terms of accuracy, precision, recall and F-score.

Existing methodology in sentiment analysis had used bag of words, parts-of-speech tagging as feature extraction techniques on text [2]. The work is limited to classify domain specific sentiment and resulted in document level sentiment classification with poor efficiency. Evolved topic modelling and by the use of LDA, it is made possible to classify sentiments by grouping into topics. But this approach is limited to automate the process of assigning labels by grouped topics, where manual assignment is needed. The literature in this paper is carried out to analyze the impact of aspect-based sentiment analysis by considering linguistic and acoustic features.

A. Linguistic Features: Aspect-based Sentiment Analysis on Text Data

Sentiment analysis had a wide variety of applications experimented on textual data. The evolution of sentiment analysis made the job of many real time applications easy in commercial markets for analyzing the customer, employee feedback in a working organization, recommending a product in ecommerce, decision making in any kind of purchase, political opinion, movie reviews, etc. Many studies have been carried to identify sentiments on text at various levels like document level, sentence level, aspect level, context level [1].

Md. E. Mowlaei et al; proposed adaptive lexicon-based ABSA [2] using three different types of lexicons like opinion lexicon, Sent-WordNet, Subjective to implement dynamic aspect-based sentiment analysis. The proposed methodology overcomes the limitations of existing domain dependent static lexicon approaches. The model lacks to identify implicit aspects even though it draws the attention to identify context dependent aspects in a dynamic way.

O. Alqaryouti et al; to improve the efficiency of sentiment classification proposed an integrated lexicon and rule-based approach [3] for aspect-based sentiment analysis to identify both implicit and explicit aspects. But the Lexicons used for generating the aspects are manually assigned to achieve higher efficiency in identifying the implicit and explicit aspects. A rule-based approach is used to integrate the extracted aspects and sentiments for classification. The model is implemented on government review data where general public post their opinions and it was suggested that it can be useful in mobile apps to analyze the feedback from public or customers.

V.S. Anoop et al; proposed an aspect-based sentiment analysis model on text using a topic modelling technique called LDA [4]. The input text by the use of LDA algorithm is segmented into topics, which then mapped manually to a relevant aspect. In case where there is a need to process huge data for sentiment analysis, it will be very difficult.

M. Shams et al; proposed a language independent aspect-based sentiment analysis model which undergoes through three phases of fine-grained operations [5]. The aspects are extracted

by having prior knowledge on dataset been used and used aspect word sets for mapping the polarity to the aspect. And finally used an expectation-maximization algorithm for calculating weightage of each word with respect to its aspect and assigned sentiment.

M. Syamala et al; to overcome the limitation of manual topic label assignment to the topics extracted from LDA proposed a deep fusion mechanism [19]. The extracted topics from LDA are converted into word embeddings and trained over a one-layer neural network to determine topic label for each set of extracted topics. The proposed sentiment classification model was compared against the models implemented with LDA and without LDA.

B. Acoustic Features: Aspect-based Sentiment Analysis on Audio Data

Most of the research carried out on speech for analysis is on either speech recognition or emotion recognition. Emotion recognition in speech differs from sentiment identification in text. Recognition of emotion in speech depends on various factors like pitch, volume, frequency, time, intensity, jitter, noise, and etc. But in case of text, identification of sentiment is independent of all the external environmental factors. So, there is a need to know the fusion mechanism between speech and text features for performing sentiment analysis. In this section, some of the existing works carried out on speech data for sentiment analysis is presented.

D. Griol et al; proposed a fusion mechanism between speech and text features [6]. The features extracted from both these modalities are trained for emotion classification in speech and sentiment classification in text. Acoustic and contextual features are been extracted from speech for emotion classification and semantic features are been extracted from transcriptions for sentiment identification. The proposed fusion model takes the account of peculiar context related errors in the transcriptions derived from speech.

Zhiyun Lu et al; proposed an end-to-end automatic speech emotion recognition model using pre-trained speech and text features from IEMOCAP dataset [7]. Build a speech sentiment database to enhance the sentiment in speech and also which is been considered as one of the current challenges in this field of research. The trained features are classified using a self-attention Recurrent Neural Network (RNN) to differentiate sentiment with respect to language model.

Bryan Li et al; combined acoustic and lexical features to develop a sentiment analysis model in order to analyse customer call services [8]. The acoustic low-level descriptors like MFCC, Intensity, pitch, loudness features are extracted using open SMILE. Lexical features are been extracted by considering n-grams. The Lexical classifier model was built on IEMOCAP dataset to make a comparison between the speech transcriptions and to choose the perfect speech recognition model. Implemented a decision-level fusion mechanism also known to be a late fusion to train the two modalities input to a classifier for decision making or classification.

Dong Zhang et al; proposed a REINFORCED approach which differs from self-attention model by concentrating on the word-level features in both speech, text and avoided the low level weighted and noisy features [9]. In this paper, the title of the paper is to depict sentiment on speech and text but actually emotions are been classified by training the extracted features into a deep learning model using SoftMax layer.

Maghilnan S et al; proposed speech sentiment analysis on speaker specific data [10]. In the proposed model, conversation between two entities is taken as input but can't able to handle if both the entities speak simultaneously. Two independent tasks are carried out to perform speaker identification and speech transcribes generation. Later both these outputs are used to map the transcribed text with respect to its speaker ID. Finally, the output text dialogue is classified into sentiment based on its polarity.

IV. Deep Fusion of Linguistic and Acoustic Features

As the proposed model in this paper analyses ABSA by considering both speech and text data, it's important to know the different ways of fusing linguistic and acoustic features. In general, the research that is been carried out in this area, defines three basic variants of fusing mechanisms like feature-level fusion, decision-level fusion, hybrid-level fusion.

A. Feature-Level Fusion

Feature-level fusion is also known as early-fusion where features from various modalities are extracted separately and a deep classification analysis was performed by fusing the models to enhance the performance. The main advantage with this type of fusion is, in the early stage it helps to derive or extract modality dependent features making the models to achieve more improvement. The main drawback with feature level fusion is that the aspects with respect to the modality may differ and accurate analysis can't be achieved when combined analysis is performed. For example, in speech the features are acoustic and in text features are linguistic. Poria S et al; in his paper multimodal emotion recognition and sentiment analysis [11], used feature-level fusion to fuse three modalities of YouTube data. A deep convolutional neural network was used to extract speech and visual features and word-embeddings, parts-of-speech tagging was used to extract textual features. A multiple kernel learning classifier is used to fuse and analyze the sentiment.

B. Decision-Level Fusion

In decision-level fusion, the features from different modalities are extracted separately and classified separately. The results obtained from each classification are merged into a feature vector for final decision making. The advantage of this approach is that the final feature vector obtained from decision fusion of individual modalities will be in same format so that no conversion is required. The drawback of this fusion is to perform classification on different modalities involves different types of classifiers.

Wöllmer M et al; used decision level fusion mechanism in his paper [12] to fuse audio, visual and text features of YouTube input data. The extracted acoustic, visual features are trained by a LSTM for sentiment score evaluation and Support Vector Machine (SVM) is used to train and derive the sentiment score of textual features. Final decision level late fusion was performed for final sentiment prediction by

calculating the weighted sum on the sentiment score obtained by assigning a weightage of about 1.2 to linguistic and 0.8 for audio and visual score.

C. Hybrid-Level Fusion

Hybrid-level fusion includes the model to use both feature and decision-level fusion mechanism in order to overcome the drawbacks in individual fusions.

Yue Gu et al, proposed an attention-based hybrid multimodal network for spoken language classification using hybrid fusion approach [13]. Word2Vec and Mel-frequency spectral coefficients (MFSCs) of text and audio features are been extracted. The extracted features are individually trained over a LSTM to obtain informative context related words and frames undergoing a feature level fusion. And finally, modality level fusion i.e., a decision-level fusion is performed by passing the extracted individual test and audio features through an attention layer to extract informative modality level features.

V. Proposed Model

In this paper, a novel Aspect-based Sentiment Analysis model was implemented on speech and text data. The dataset used for implementing the model is drawn from YouTube social platform. In order to evaluate the experimental results for sentiment modality comparison, both the speech and text models are been tested on the same dataset. The domain chosen for carrying out our experimental analysis is real-time product review data. In the initial phase, the raw audio format of the product review YouTube Video is trained over a speech analysis model. The speech analysis model maps the acoustic spectrogram features of the speech signal into the respective word utterances using a deep learning and language model. The word utterances from the speech analysis model are trained over different variants of text feature extraction techniques for deriving related and relevant aspects. The sentiment with respect to the derived aspect is analyzed for performing Aspect-based sentiment analysis. The components (features) in speech and text data are processed individually and are then fused. So, the whole process uses a hybrid fusion mechanism for mapping speech and text features for performing ABSA. Fig. 2 explains the work flow of the proposed speech and text analysis model for efficient Aspect based Sentiment Analysis.

A. YouTube Product Review Data Collection and Processing

In this phase, YouTube product reviews of Samsung M31mobile were downloaded as dataset. YouTube, a social platform where people share their live experience in the form of reviews have a natural, spontaneous speaking style. As the way the speaker speaks have a direct impact on describing the accuracy of the model, made me to motivate and download the dataset from YouTube for performing Aspect-based sentiment analysis on speech data. In total 40 YouTube reviews of size 90 KB on the Samsung M31 product having strong presence of subjectivity, positivity and negativity are randomly collected and converted to .wav files, are used for ABSSA.

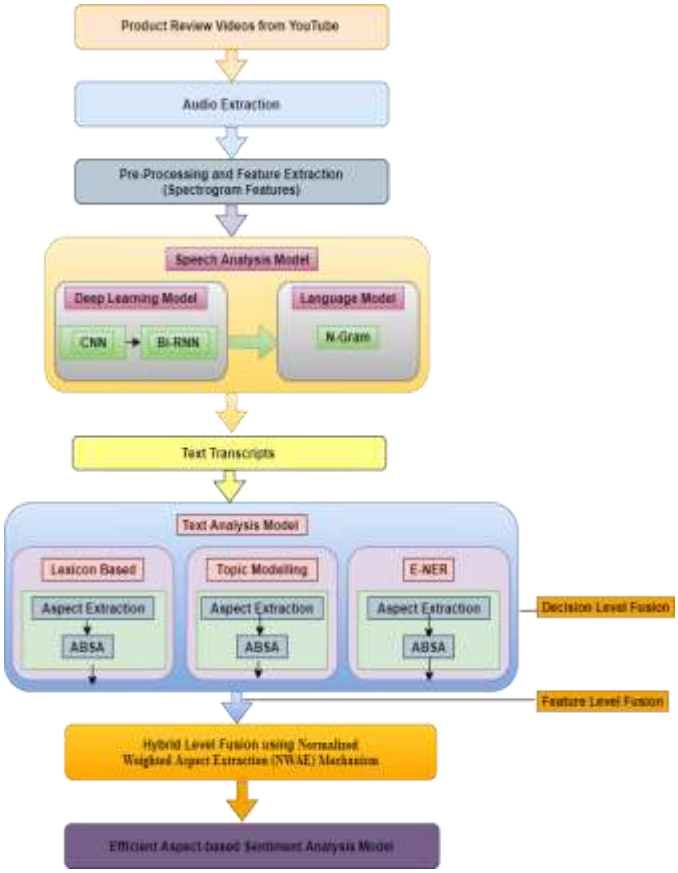

Fig. 2. Work flow of Proposed Speech and Text Analysis Model for Efficient Aspect-based Sentiment Analysis.

B. Speech Analysis Model

Emotion in speech is treated to be a kind of sentiment, which expresses an individual feeling in terms of happy, sad, fear, disgust, angry and etc. Sentiment in text differs from emotion in speech and there is a need to perform speech analysis in the form of automatic speech recognition to enhance sentiment from audio data. There are many ASR models and online speech - text conversion API's.

To enhance the performance of traditional ASR models and to overcome the limitations in online speech-text API's, in our proposed model used a deep learning framework for analyzing the acoustic features and a bi-gram language model to map the word utterances. As sentiment analysis is independent on the speech features like pitch, intensity, volume and etc. Initially, the spectrogram features are extracted from the input Wav audio file and are trained over a Convolutional Neural Network and a Bi-directional Recurrent Neural Network (Bi-RNN). The acoustic features when trained over these deep neural networks produces a character sequence of spoken utterances. A bi-gram language model by the use of chain rule retrieves the maximum occurrence of character sequences and the same are mapped into word utterances. Fig. 3 shows the text transcripts extracted from the proposed speech analysis model.

Source.Name	message
Samsung Galaxy M21 review _ Better than the Ga...	okay so what are the questions that I get ask...
Samsung Galaxy M31 - Full Review, 64 MP Quad C...	so after the galaxy m31 the m30 s we now have ...
Samsung Galaxy M31 Detailed Camera Review - En...	hey guys did saga from tech works and in this...
Samsung Galaxy M31 Full Review _ Better than R...	hey guys this is one from guiding dick and to...
Samsung Galaxy M31 Review - Decent Device, Dis...	some sense m-series our phones have always bee...
Samsung Galaxy M31 Review 2 Weeks Later! The B...	what's going on guys my name is Wade with tech...
Samsung Galaxy M31 Review with Pros.txt	hi guys this is Ranji them in this video let's...
Samsung Galaxy M31 Review_ Should You Buy_ - E...	so the galaxy m31 arrived in India recently a...
Samsung Galaxy M31 review_ Worth all the Hype_...	Samson's MCV smartphones have been quite popul...
Samsung Galaxy M31 The Only Review You Need - ...	you guys welcome to I again today we're going...

Fig. 3. Extracted Text Transcripts from the Proposed Speech Analysis Model.

1) Creation of Spectrogram

Input: Audio signal
Output: One-Time frame vector
Step 1: Dividing the input audio signal into time frames of frame size 1024, with a sampling rate of 16 kHz.
Step 2: Each frame signal is then split into its frequency components with a hop size of 512 samples between each successive Fast Fourier Transform window.
Step 3: Finally, each time frame is then represented as a one-time frame vector with a vector of amplitudes at each frequency.
Step 4: The one-time frame vectors obtained when lined up in time series order gives us the visual representation of input audio signal as a spectrogram.

2) Language model: The word sequences obtained from the above acoustic model need to be refined as the acoustic model finds the probability of character utterances based on sound and there are cases where two words can utter same sound. The use of language modelling followed by acoustic modelling helps to rectify this problem and increases the likelihood score of a particular sequence of word utterances. Equation (1) formulates the representation of a sequence of word utterances using N-gram language model. In this proposed speech analysis model, used a bi-gram language model by a chain of rule mechanism to find the respective sequence probability.

$$P(\mathrm{Word}_1, \mathrm{Word}_2, \ldots . \mathrm{Word}_n) = \pi\, \mathrm{P}(\mathrm{Word}_i | \mathrm{Word}_1, \mathrm{Word}_2, \ldots\, \mathrm{Word}_{i-1}) \quad (1)$$

3) Text analysis model: For improving the efficiency of Aspect-based sentiment analysis, in this paper three variants of text feature extraction techniques are applied for aspect level feature extraction. Sentiment was analyzed with respect to the extracted aspect at decision level. The way the decision level aspects are extracted and analyzed for sentiment are presented in detail in the below section.

a) Lexicon based semi-supervised pattern generation technique (Model 1): The opinion words representing the sentiment called as aspect terms are extracted by generating patterns from bi and tri grams. As defined, it is a Lexicon based approach, the similar aspect words related to the input dataset are assigned statistically. In addition to statistically stuffed aspect words, the hypernyms of aspect terms extracted from patterns are generated by using wordnet. The final aspect terms are obtained by considering statistically assigned words and its similar words generated from the patterns.

Consider a set of word sequences

$$\mathrm{Word}_1{}^n = \mathrm{Word}_1 \ldots\ldots\ldots \mathrm{Word}_n \quad (2)$$

Bigram approximation is represented as

$$\mathrm{P}(\mathrm{Word}_1{}^n) = \prod_{k=1}^{n} \mathrm{P}(\mathrm{Word}_k / \mathrm{Word}_{k-1}) \quad (3)$$

N-gram approximation is represented as

$$\mathrm{P}(\mathrm{Word}_1{}^n) = \prod_{k=1}^{n} \mathrm{P}(\mathrm{Word}_k / \mathrm{Word}_{k-N+1}^{k-1}) \quad (4)$$

The final aspect terms obtained are mapped with the patterns generated to extract the sentiment terms. Sentiment score was computed on the trained aspect and sentiment terms by importing 'testimonial. sentiment. polarity' library.

b) Topic Modelling Technique LDA (Latent Dirichlet Allocation) (Model 2): Text features are been extracted from the pe-processed input using LDA. Python libraries like gensim and ldamallet are used for extracting the dominant topics (aspect terms). Extraction is done by calculating the term document frequency on pre-processed lemmatized data by considering NOUN, ADJ, ADV and VERB from n-gram data. Dominant topic words termed to be as aspects with respect to the input qualifying the sentiment are extracted. Some of the examples of aspect terms in the context of electronic gadgets are battery, display, power etc.

Probability based topic extraction using LDA is formulated in (5).

$$\mathrm{P}(z = t/w) \infty \left(\alpha_t + n_{t/d}\right) \frac{\beta + n_{w/t}}{\beta V + n_{\cdot/t}} \quad (5)$$

Aspect category groups a list of aspect terms into its relevant category. For example, the aspect terms like battery, display, power can be categorized under the category mobiles and similarly taste, flavor, ambience can be categorized under the category restaurant. Aspect category is detected by training the extracted dominant/aspect terms into a Convolutional Neural Network (CNN). Using polarity as a measure, respective sentiment terms are extracted from the extracted aspect category terms.

c) Efficient Named Entity Recognition (E-NER) (Model 3): The aspect terms in this approach are extracted by a dependency parsing mechanism using POS tagging, an NLP technique.

A convolutional neural network was used to map the extracted aspect as relevant aspect categories. Word embeddings mechanism (6), (7) is used to train the input aspect terms as vectors to CNN (8). Filtering aspect related sentiment words and aspect sentiment classification uses the same

methodology as followed for aspect category detection for aspect term polarity extraction and sentiment classification.

$$m_i = \sum_{j=1,j\neq i}^{n}(a_{ij}, w_j) \quad (6)$$

$$a_{i,j} = \frac{\exp(score(w_i,w_j))}{\sum_{j1=1}^{n}\exp(score(w_i,w_j))} \quad (7)$$

$$score(w_i, w_j) = v_a^T \tanh(W_a[w_i \oplus w_j]) \quad (8)$$

4) Hybrid level fusion: In text analysis model, by the three variants of aspect extraction techniques we performed decision level fusion for analyzing the sentiment. The extracted aspects with respect to their sentiment have undergone a feature level fusion for enhancing the performance. By means of this decision level fusion followed by feature level fusion, helps to overcome the problems of dimensionality and filters the weighted aspects by deriving improved performance. In hybrid level fusion phase, employed a Normalized Weighted Aspect Extraction (NWAE) mechanism (10) in which the aspects extracted from each technique are filtered based on their weights. A decision rule was applied for classifying the polarity class of the derived weighted aspects (11).

$$tf - idf(d_i, v_j) = \frac{count(d_i v_j)}{\sum_{v^1 \in d_i} count(d_i, v^1)} \times \log \frac{n}{|\{v_j \in d^1, d^1 \in D\}|} \quad (9)$$

$$NWAE = \frac{1}{|P_j|}\sum_{d_i \in p_j} d_i \quad (10)$$

$$y_t = arg\ max_{pj} \sum_{x_i \in x_t}(1 - dist(x_t, x_i))I(x_i, c_j) \quad (11)$$

Input: Extracted Aspect terms A_t in d_1, d_2, d_3.
Output: Weighted Aspects W_a and its polarity.

```
file_path <- file.path(" d_1, d_2, d_3")
docs <- Corpus (DirSource (file_path))
docs <- Corpus (VectorSource(docs)) #This tells R to treat your preprocessed documents as text documents.
dtm <- DocumentTermMatrix(docs)
tdm <- TermDocumentMatrix(docs)
dtms <- removeSparseTerms(dtm, 0.1) #  Start by removing sparse terms
tf.idf<- weightTfIdf(dtm,normalize=TRUE)
x <- apply (tf.idf, 1, sum) #computing sum of the rows in tf-idf matrix
d <- NULL
for (i in seq(docs)) #NWAE
  d[i]<-x[i]/2
#Sum of the squares of the NWAE
df <- NULL, df1 <- NULL, df1 <-0
for (i in seq(docs))
  df[i] <- (d[i]) ^2
  df1<- df1+df[i]
#Sum of the squares of the documents
n <- ncol(dtm), n1 <- nrow(dtm), sf <- NULL, s <-NULL, w <- NULL
for (z in seq(docs))
  sf[z] <- 0
  for (i in seq(n))
    w[i]<- inspect (tf.idf[z [1], i])
    w[i]
    s[i] <- (w[i]) ^2
    sf[z] <- sf[z]+s[i]
#Similarity function for dimensionality reduction
sim <- NULL
for (i in seq(docs))
  sim[i] <- (x[i]*d[i])/((sqrt(sf[i])) * (sqrt(df1)))
#Projected documents values sum of the squares
p <- NULL, pro <- NULL, p <- 0
for (i in seq(docs))
  pro[i] <- (sim[i]) ^2
  p <- p+pro[i]
#Normalize the projected document vectors
v <- NULL
for (i in seq(n1-1))
  v[i] <- (sim[i]*sim[n1])/((sqrt(p)) * (sqrt((sim[n1]) ^2)))
#Compute Euclidean distance
m <- NULL
for (i in seq(n1-1))
 m[i]<-dist(v[i], sim[n1])
#Apply decision rule
y <- NULL
for (I in seq(n1-1))
 y[i] <- (1-m[i])
 max(y)
for (i in seq(n1-1))
 if(y[i]==max(y))
   print ("The aspect term is classified as:")
   print(i)
```

VI. Experimental Results

Experimental results in this paper are been evaluated on product review content drawn from YouTube videos. The data set in its initial video format is processed into Wav files for performing speech recognition using deep learning and language models. Word Error Rate (WER) and Character Error Rate (CER) are the two-evaluation metrics used for recognizing the performance of the speech recognition model. The proposed speech recognition model proved to improve the efficiency of the model by achieving 5.7% WER and 3% CER. The results proved to improve the performance of the proposed model when compared with the traditional state of art methods. The aim to perform aspect-based sentiment analysis on speech data made us to carry out the analysis on the derived text transcripts by using three different feature extraction techniques.

The application of feature extraction techniques improved the efficiency of the proposed model. The fusion of aspects at decision-level after undergoing individual feature-level fusion improvises the feature selection process and overcomes the problem of dimensionality.

Experimental results obtained from the three different text analysis models, discussed in Section 5 are made a comparison. Accuracy, precision, recall and f-score are the metrics used to measure the performance of the proposed model.

Fig. 4 lists the different aspects extracted from the patterns generated by means of bi-gram and tri-grams in text analysis model1.

Fig. 4. List of Aspects Extracted from Model 1.

Model 1, uses a semi-supervised pattern generation technique for aspect extraction, where it needs a list of statistically assigned aspects (Lexicon) relevant to the taken input context. So, Fig. 5 shows the list of statistically assigned aspects used in Model 1.

```
stuff = ['software', 'application', 'service', 'power supply', 'sim card', 'display',
         'storage space', 'sensor', 'wireless charging', 'design', 'cpu', 'accessories',
         'camera','quality','time','condition','screen','price','case','build','access',
         'battery','buy','power','switch','light','design','technology','radio','fashion'
'product','charging','feature','touch','profile','car','slot','tables','construction',
'period ','system','game','bottom','sound','blackberry charge','price anyone','price extra',
'cord length','charge port',' phone','horizon charge','fraction price','charge ','key',
'extension','internet','cheap','cover','speaker']
```

Fig. 5. Statistically Assigned Aspects in Model 1.

For effective aspect extraction in model 1, hypernyms are generated for the statistically assigned aspects. Extracted hypernyms for the statistically stuffed aspects are shown in the below Fig. 6.

```
Meaning 0 NLTK ID:  software.n.01
Hypernyms: code, computer_code
Hyponyms: alpha_software, authoring_language, beta_software, compatible_software, computer-aided_design,
CAD, database_management_system, DBMS, freeware, groupware, operating_system, OS, program, programme, computer_program, compu
ter_programme, routine, subroutine, subprogram, procedure, function, shareware, shrink-wrapped_software, software_documentati
on, documentation, spyware, supervisory_software, upgrade

Meaning 0 NLTK ID:  application.n.01
Hypernyms: use, usage, utilization, utilisation, employment, exercise
Hyponyms: misapplication, technology, engineering

Meaning 1 NLTK ID:  application.n.02
Hypernyms: request, petition, postulation
Hyponyms: credit_application, job_application, loan_application, patent_application
```

Fig. 6. Extracted Hypernyms for the Statistically Stuffed Aspects in Model1.

Table I and Fig. 7 shows the performance analysis comparison of model 1 when validated using machine learning algorithms in terms of accuracy, precision, recall and f1-score. From the analysis it shows that decision tree algorithm derived better accuracy of 73% among all the other compared machine learning algorithms.

TABLE I. PERFORMANCE ANALYSIS OF MODEL 1 USING MACHINE LEARNING ALGORITHMS

Machine Learning Algorithm	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Support Vector Machine	65	65	65	65
Naïve Bayes	65	56	65	55
Logistic Regression	65	63	65	75
Decision Tree	**73**	**72**	**73**	**71**
K-Nearest Neighbor	67	44	67	53
Random Forest	69	68	69	68

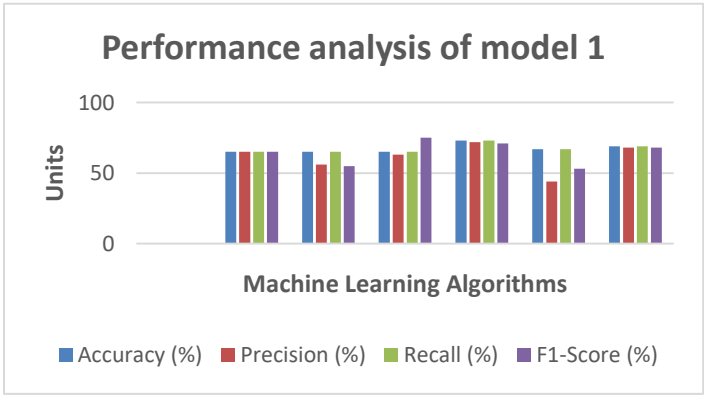

Fig. 7. Performance Analysis of Model 1 using Machine Learning Algorithms.

Fig. 8 presents the list of aspects extracted in model 2 based on probability using python libraries like gensim and ldamallet.

Fig. 9 presents the extracted aspect sentiment classification and plot in Fig. 10 presents the top most salient features extracted using model 2.

```
[(15, [('time', 0.12234042553191489), ('back', 0.0851063829787234), ('angle', 0.06914893617021277),
rocessor', 0.047872340425531915), ('single', 0.03723404255319149), ('bit', 0.03723404255319149), ('
ung', 0.02307692307692231), ('full', 0.06666666666666667), ('run', 0.06153846153846154), ('amole',
1025641025641026), ('notification', 0.041025641025641026), ('panel', 0.041025641025641026), ('hd',
1610486891385768), ('range', 0.11235955056179775), ('mode', 0.08239700374531835), ('detail', 0.0636
247191), ('update', 0.0299625468164794), ('hdr', 0.026217228464419477), ('light', 0.026217228464419
3659), ('photo', 0.07317073170731707), ('phone', 0.04065040650406504), ('performance', 0.036585365
('question', 0.032520325203252036), ('shoot', 0.032520325203252036), ('comparison', 0.032520325203
1527), ('make', 0.10843373493975904), ('bit', 0.08032128514056225), ('picture', 0.06827309236947791
o', 0.040160642570281124), ('portrait', 0.040160642570281124), ('expect', 0.040160642570281124), ('
y', 0.1188118811881188), ('quality', 0.07425742574257425), ('test', 0.0594059405940594), ('speaker'
5445544554455), ('pro', 0.03960396039604), ('lag', 0.034653465346534656), ('user', 0.024752475
6108594), ('series', 0.04524886877828054), ('offer', 0.04524886877828054), ('year', 0.040723981900
('line', 0.027149321266968326), ('network', 0.027149321266968326), ('cheap', 0.02262443438914027)])
('shot', 0.10989010989010989), ('ultra', 0.06227106227106227), ('focus', 0.047619047619047616), ('b
2197802197802198), ('turn', 0.02197802197802198), ('noise', 0.018315018315018316)]), (1, [('wide',
3506493), ('primary', 0.06060606060606061), ('side', 0.05627705627705628), ('great', 0.051948051948
('manage', 0.025974025974025976), ('switch', 0.025974025974025976)]), (17, [('thing', 0.14794520547
('work', 0.05205479452054795), ('guy', 0.05205479452054795), ('feel', 0.038356164383561646), ('he
136986301369864), ('mid', 0.024657534246575342)])]
```

Fig. 8. List of Aspects Extracted based on Probability from Model 2.

Dominant_Topic	Topic_Perc_Contrib	Keywords	Text	sentiment_terms	pol	sentiment
0.0	0.5640	phone, video, display, price, time, pretty, gu...	okay so what are the questions that I get ask...	ask awful good vary late break well launch pla...	0.9961	positive
1.0	0.6080	good, camera, battery, thing, game, screen, ta...	so after the galaxy m31 the m30 s we now have ...	s come adaptive fast charge case minimal bare...	0.9993	positive
1.0	0.8563	good, camera, battery, thing, game, screen, ta...	hey guys did saga from tech works and in this...	go detailed samsung current get take usual req...	0.9260	positive
0.0	0.6546	phone, video, display, price, time, pretty, gu...	hey guys this is one from guiding dick and to...	guide review recall watch check long think rig...	0.9882	positive
1.0	0.6116	good, camera, battery, thing, game, screen, ta...	some sense m-series our phones have always bee...	great amole create average bring let find know...	0.9986	positive
0.0	0.7038	phone, video, display, price, time, pretty, gu...	what's going on guys my name is Wade with tech...	go samsung vast base experience strike awesome...	0.9993	positive
1.0	0.8204	good, camera, battery, thing, game, screen, ta...	hi guys this is Ranji them in this video let's...	let test share feel divide quick big come mass...	0.9986	positive
1.0	0.6218	good, camera, battery, thing, game, screen, ta...	so the galaxy m31 arrived in India recently a...	arrive certain game compare buy galaxy let ans...	0.9991	positive
0.0	0.5300	phone, video, display, price, time, pretty, gu...	Samsoni's MCV smartphones have been quite popul...	popular different desirable hype mid range die...	0.9990	positive
0.0	0.6112	phone, video, display, price, time, pretty, gu...	you guys welcome to I again today we're going...	welcome go review new samsung launch start gre...	0.9990	positive
1.0	0.5094	good, camera, battery, thing, game, screen, ta...	Phone bckup is not good	good	0.4404	negative

Fig. 9. Aspect based Sentiment Classification in Model 2.

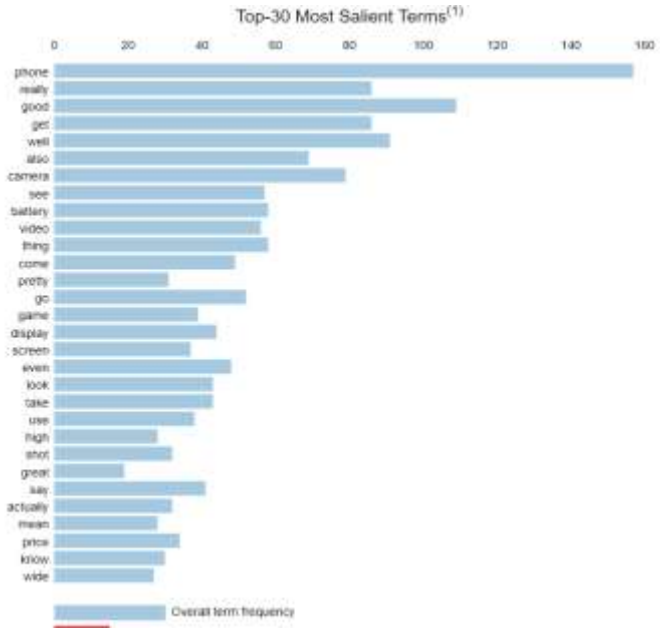

Fig. 10. Most Salient Aspects Extracted from Model 2.

Table II and Fig. 11 shows the performance analysis comparison of model 2 when validated using machine learning algorithms in terms of accuracy, precision, recall and f1-score. From the analysis it shows that Random Forest algorithm derived better accuracy of 89% among all the other compared machine learning algorithms.

TABLE II. PERFORMANCE ANALYSIS OF MODEL 2 USING MACHINE LEARNING ALGORITHMS

Machine Learning Algorithm	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Support Vector Machine	74	69	67	64
Naïve Bayes	50	83	50	50
Logistic Regression	74	72	67	65
Decision Tree	75	88	75	77
K-Nearest Neighbor	75	56	75	64
Random Forest	89	87	85	85

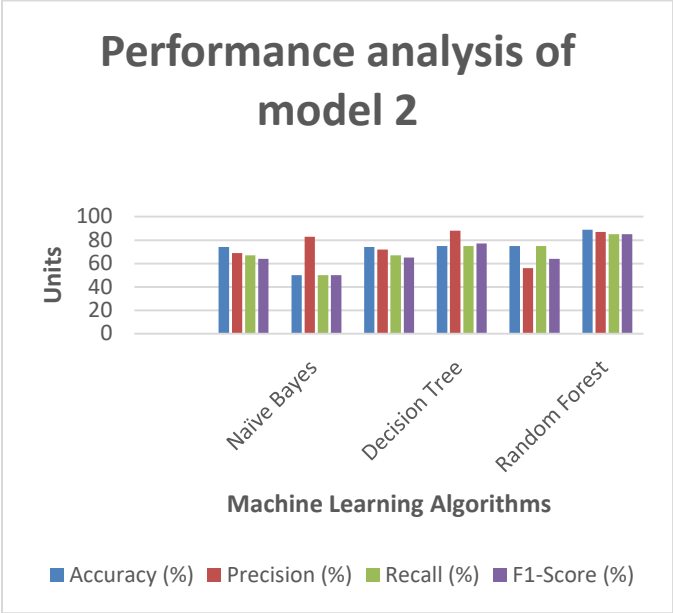

Fig. 11. Performance Analysis of Model 2 using Machine Learning Algorithms.

The Fig. 12 presents the way the aspects are extracted using POS tagging by dependency parsing mechanism and Fig. 13 presents the aspect-based sentiment analysis on the derived aspects using model 3.

Table III and Fig. 14 shows the performance analysis comparison of model 3 when validated using machine learning algorithms in terms of accuracy, precision, recall and f1-score. From the analysis it shows that Random Forest algorithm derived better accuracy of 95% among all the other compared machine learning algorithms.

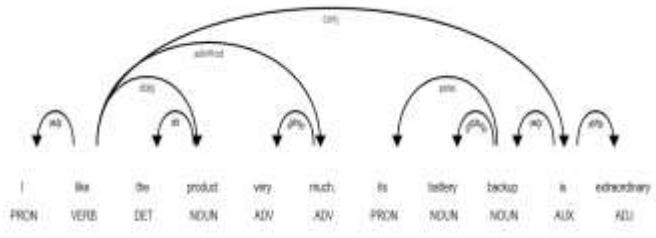

Fig. 12. Aspects Extracted from Model3.

Source_Name	message	Column2	aspect_terms	aspect_category	sentiment_terms	pol	sentiment
Samsung Galaxy M21 review _ Better than the Ga...	okay so what are the questions that I get ask...	100 and for the price this is the best Samsung	questions lot smartphone rupees answer time bi...	rupees smartphone n21 smudges buttons consumpt...	ask awful good vary late break well launch pla...	0.9981	positive
Samsung Galaxy M31 - Full Review, 64 MP Quad C...	so after the galaxy m31 the m30 s we now have ...	NaN	m30 m31 variants gigs ram storage box charger ...	phones case example smartphones issues sides g...	s come adaptive fast charge case minimal bare ...	0.9993	positive
Samsung Galaxy M31 Detailed Camera Review - En...	hey guys did saga from tech works and in this ...	000 rupees these have all been 16 megapixel im...	guys works video look cameras phone days silua...	guys samples sensor sensor sensor bones modes ...	go detailed samsung current get take usual req...	0.9260	positive
Samsung Galaxy M31 Full Review _ Better than R...	hey guys this is one from guiding dick and to...	NaN	guys guys video device link box device time mo...	guys device guys things body terms lights imag...	guide review recall watch check long think rig ...	0.9862	positive
Samsung Galaxy M31 Review - Decent Device, Dis...	some sense m-series our phones have always bee...	000 milliamp hours plus 15 watt fast charging ...	phones option displays life cameras m31 table ...	option levels flowers colors details booker ca...	great amoled create average bring let find know...	0.9985	positive
Samsung Galaxy M21 Review 2 Weeks Later! The B...	what's going on guys my name is Wade with tech...	NaN	guys name weeks majority youtubers phone peopl...	guys people m21 folks opinion support thing co...	go samsung vast base experience strike awesome ...	0.9993	positive
Samsung Galaxy M31 Review with Pros,txt	hi guys this is Ranji them in this video let's ...	000 anyways guys that's it for now for the rev...	guys video review guys device week experience ...	guys thing colors thing phones fingers device ...	let test share feel divide quick big come mass...	0.9886	positive
Samsung Galaxy M31 Review_ Should You Buy_ - E...	so the galaxy m31 arrived in India recently a ...	NaN	questions phone performance cameras review que...	design colors movies phones details photos thi...	arrive certain game compare buy galaxy let ans...	0.9991	positive
Samsung Galaxy M31 review_ Worth all the Hype_ ...	Samson's MCV smartphones have been quite popul...	NaN	smartphones segments centric qualities phone y...	smartphones qualities guys guys list changes m...	popular different desirable hype mid range die...	0.9990	positive
Samsung Galaxy M21 The Only Review You Need - ...	you guys welcome to I again today we're going ...	NaN	guys price features phone rupees line sale 5th...	price set specifications display display displ...	welcome go review new samsung launch start gre...	0.9990	positive
Samsung Galaxy M31 coder review	Phone bckup is not good	NaN	bckup		good	0.4404	negative
Samsung Galaxy M31 coder friend review	sleepy phone	NaN	phone	phone	sleepy	0.0000	negative

Fig. 13. Final Output of Aspect-based Sentiment Analysis.

TABLE III. PERFORMANCE ANALYSIS OF MODEL 2 USING MACHINE LEARNING ALGORITHMS

Machine Learning Algorithm	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Support Vector Machine	75	72	69	65
Naïve Bayes	50	83	67	64
Logistic Regression	74	74	72	72
Decision Tree	50	77	65	72
K-Nearest Neighbor	72	64	62	60
Random Forest	95	92	90	90

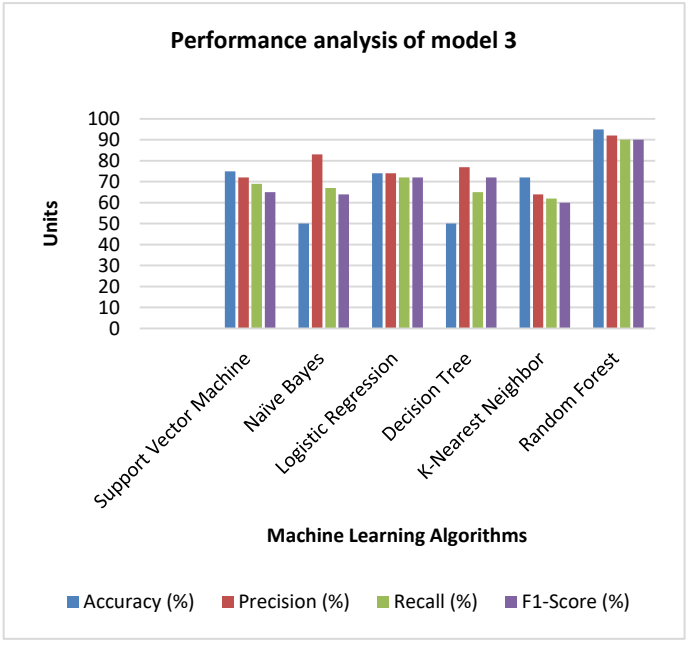

Fig. 14. Performance Analysis of Model 3 using Machine Learning Algorithms.

VII. CONCLUSION

The motivation to achieve fine-grained aspect-based sentiment analysis made us to propose an efficient hybrid aspect-based sentiment analysis model by the fusion of speech and text aspects. To enhance the performance of traditional ASR models, a deep learning framework and a bi-gram language model is employed for deriving the speech-to-text aspects. The results are made a comparison with the traditional ASR models and the proposed model achieved 5.7% WER and 3% CER [17]. Three variants of text feature extraction techniques were employed for improving the efficiency of ABSA. A decision level fusion was performed on the aspects extracted to enhance the sentiment in an efficient way. Feature level fusion was applied by proposing a NWAE mechanism as a feature selection measure to overcome the problem of dimensionality. The obtained results obtained from speech and three variants of text analysis models are compared with the individual text feature extraction techniques [18] [19] [20] and proved that the proposed hybrid level fusion mechanism improves the user readability by faster access and there by improves the performance.

REFERENCES

[1] M. Syamala, and N.J.Nalini, "A Deep Analysis on Aspect based Sentiment Text Classification Approaches," International Journal of Advanced Trends in Computer Science and Engineering, vol. 8, No.5, pp.1795-1801 ,September - October 2019.

[2] Md. E. Mowlaei, Md. S. Abadeh, and H. Keshavarz, "Aspect-based sentiment analysis using adaptive aspect-based lexicons," Expert Systems with Applications, vol. 148, pp.1-27 ,15 June 2020.

[3] O. Alqaryouti, N. Siyam, A.A. Monem, and K. Shaalan, "Aspect-Based Sentiment Analysis Using Smart Government Review Data," Applied Computing and Informatics, pp.1-13, November 2019.

[4] V.S. Anoop and S. Asharaf, "Aspect-Oriented Sentiment Analysis: A Topic Modeling-Powered Approach," J. Intell. Syst., vol. 29(1), pp.1166–1178, December 2018.

[5] M. Shams, N. Khoshavi, and A. Baraani-Dastjerdi, "LISA: Language-Independent Method for Aspect-Based Sentiment Analysis," IEEE Access, vol.8, pp. 31034-31044, February 2020.

[6] D. Griol, J. M. Molina, and Z. Callejas, "combining speech-based and linguistic classifiers to recognize emotion in user spoken utterances," Neurocomputing, vol. 326-327 ,pp. 132-140, January 2019.

[7] Z. Lu, L. Cao, Y. Zhang, Ch.Ch. Chiu, and James Fan, "Speech Sentiment Analysis Via Pre-Trained Features from End-To-End ASR Models," ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Barcelona, Spain, Spain, pp. 7149-7153, May 2020.

[8] B. Li, D. Dimitriadis, and A. Stolcke, "Acoustic and Lexical Sentiment Analysis for Customer Service Calls," ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, United Kingdom, United Kingdom, pp. 5876, 5880, May 2019.

[9] D. Zhang, S. Li, Q. Zhu, and G. Zhou, "Effective Sentiment-relevant Word Selection for Multi-modal Sentiment Analysis in Spoken Language," Proceedings of the 27th ACM International Conference on Multimedia, pp.148–156, October 2019.

[10] S. Maghilnan, and M. Rajesh Kumar, "Sentiment analysis on speaker specific speech data," International Conference on Intelligent Computing and Control (I2C2), Coimbatore, India, pp. 1-5, February 2018.

[11] S. Poria, I. Chaturvedi, E. Cambria, and A. Hussain, "Convolutional MKL based multimodal emotion recognition and sentiment analysis In Data Mining", IEEE 16th International Conference on Data Mining (ICDM) , pp. 439–448, December 2016.

[12] M. Wöllmer, F. Weninger, T. Knaup, B. Schuller, C. Sun C, K. Sagae, and LP. Morency, "Youtube movie reviews: Sentiment analysis in an audio-visual context," IEEE Intelligent Systems, vol. 28(3), pp. 46–53, March 2013.

[13] G. Yue, Y. Kangning, F. Shiyu, Ch. Shuhong, L. Xinyu and M. Ivan, "Hybrid Attention based Multimodal Network for Spoken Language Classification," Proceedings of the 27th International Conference on Computational Linguistics, pp. 2379–2390, August 20-26, 2018.

[14] S. Govindaraj and K. Gopalakrishnan, "Intensified Sentiment Analysis of Customer Product Reviews Using Acoustic and Textual Features," ETRI Journal, vol. 38 (3), pp. 494-501, 2016.

[15] L. Kaushik, A. Sangwan, and J. H. L. Hansen,"Sentiment extraction from natural audio streams," In Proceedings of International Conference on Acoustics, Speech and Signal Processing, pp. 8485-8489, 2013.

[16] H.H. Do, P.W.C. Prasad, A. Maag, and A. Alsadoon, "Deep learning for aspect-based sentiment analysis: a comparative review", Expert Systems with Applications, vol.118, pp.272-299, 2019.

[17] M. Syamala, and N.J. Nalini, "A Speech-based Sentiment Analysis using Combined Deep Learning and Language Model on Real-Time Product Review", International Journal of Engineering Trends and Technology, vol. 69 (1), pp. 172-178, January 2021.

[18] M. Syamala, and N.J. Nalini, "A filter-based improved decision tree sentiment classification model for real-time amazon product review data," International Journal of Intelligent Engineering and Systems, vol.13(1), pp. 191-202, January2020.

[19] M.Syamala, and N.J.Nalini, "LDA and Deep Learning: A Combined Approach for Feature Extraction and Sentiment Analysis," 10th ICCCNT, vol. 45670, pp. 1-5, December 2019.

[20] M. Syamala, and N.J. Nalini, "ABSA: Computational Measurement Analysis Approach for Prognosticated Aspect Extraction System", TEM JOURNAL - Technology, Education, vol. 10(1), pp. 82–94, February 2021.