ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

*(A High Impact Factor, Monthly, Peer Reviewed Journal)*

*Website: [www.ijircce.com](http://www.ijircce.com)*

Vol. 7, Issue 2, February 2019

# Sentiment Analysis of Tourist Review using Supervised Long Short Term Memory Deep Learning Approach

Priyanka Thakur, Dr. Rajiv Shrivastava

M.Tech Scholar, Department of CSE, SIRT-Excellence, Madhya Pradesh, India

Professor & Director, Department of CSE, SIRT-Excellence, Madhya Pradesh, India

**ABSTRACT:** Sentiment analysis research area is a part of opinion mining in which emotional information are extracted from the texts. In this paper the task of sentiment analysis on text data is proposed. The sentiment expression is used to classify the polarity of the text review on a scale of negative to positive using Long Short-Term Memory (LSTM) deep learning algorithm. This paper is based on lexicon features of the text reviews.The research methodology is performed in four steps. In the first step lexicons are extracted from the review and their Part-Of-Speech (POS) are tagged. In the second step the extracted part of speech tags occurrence are extracted. In third step, each POS tag's sentiment score s calculated by referring to the SentiWordNet (SWN) dictionary. Subsequently, in the last step, the extracted features are fed into LSTM deep learning classifier to classify the review into either positive and negative type of review.For the result analysis, the dataset is prepared by collecting reviews from tripadvisor website and performance is analyzed using classifiers such as support vector machine (SVM), k nearest neighbor (KNN) and LSTM deep network. The simulation gives output (accuracy) for SVM about 84%, k-NN about 90% and LSTM about 98% respectively. This represents that LSTM classifier gives best result to classify reviews into positive and negative review with POS tagged lexicon features.

**KEYWORDS:** Lexicon-based approach, Sentiment analysis, SentiWordNet, Machine Learning, LSTM.

## I. INTRODUCTION

As it is known that internet is one of the most common platforms for communication among opinions of an individuals. This platform explores different fields to express anyone's review, opinions or emotions. One of the most common medium to express emotions/opinions is through reviews. Either it is in the field of movie, news, products, tweets, etc. In each and every field through review anyone can express their sentiments. As most of the people are active online for their most of the time and express their sentiments. So, day by day database of these reviews are increasing and becoming heavy. As per recent survey, there are approx. 2.4 billion active online users, who write and read online around the world [1].

Emotions are an important aspect in the interaction and communication between individuals. The exchange of emotions through text messages and posts of personal blogs poses the informal kind of writing challenge for researches. Extraction of emotions from text will applied for deciding the human computer interaction that governs communication and many additional [2]-[5]. Emotions is also expressed by a person's speech, facial and text primarily based emotion respectively. Emotions are also expressed by one word or a bunch of words. Sentence level emotion detection technique plays a vital role to trace emotions or to look out the cues for generating such emotions. Sentences are the essential info units of any document. For that reason, the document level feeling detection technique depends on the feeling expressed by the individual sentences of that document that in turn depends on the emotions expressed by the individual words. The sentiments from various texts and classifies them accordingly into positive, negative or neutral classes [6].

The analysis of feelings is an innovation that will be very important in the coming years. With Opinion Mining, you can get high quality content. An important part of the early research in this area was based on product evaluations [7] and defined feelings as positive, negative or neutral. Most opinion polls currently focus on online network sources, such as

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

**International Journal of Innovative Research in Computer and Communication Engineering**

*(A High Impact Factor, Monthly, Peer Reviewed Journal)*

*Website: www.ijircce.com*

**Vol. 7, Issue 2, February 2019**

IMDB, Twitter [8] and Facebook, and require that the methods be adapted to meet the growing interest of the public. Furthermore, the analysis of phrases classification at sentence level proves to be a difficult task.

## II. Related Work

English is the most popular language for research in Natural Language Processing. Most approaches used in this area are:

- Subjective Lexicon
- Machine Learning

### A. *Subjective Lexicon Approach*

Lexicon approach depends on finding opinion lexicon which analyzes sentiment of text. This approach has 2 methods:- Dictionary based and Corpus based. There are 3 main approaches in finding opinion list. Manual approach is very time consuming so it is combined with either of these two [9].

There are three popular methods for generation of subjective lexicon:-

- Use of Bi-lingual dictionary[2]
- Machine Translation[2]
- Use of Wordnet[4]

### B. *Machine Learning Approach*

In such way total feature vector is generated for each review using features. These features are further classified by using classifiers. For each extracted features of review emotion classification algorithm is applied on different set of inputs. Different classifiers are such as SVM, Neural Network, KNN, Random Forest etc.

Some of the contribution in this field are discussed below:

Sruthi S et al. [1] propose an entity recognition method in the preprocessing stage to eliminate the irrelevant information from the reviews. Performance of SVM is about 95% and proposed sentiment analysis system efficient in terms of time and cost. More features are required based on context.

Tirath et al. [2] extracted features which are strongly effective in deciding the extremity of the movie reviews and used computation linguistic methods preprocessing of the information. Six classification techniques are analyzed on this technique. Found that Random Forest outperforms an accuracy of 88.95%. NLP based feature extraction is not properly discussed.

Md Shad Akhtar et al. [3] proposed a novel hybrid deep learning architecture which is highly efficient for sentiment analysis. The selected features are optimized by selecting through a multi-objective optimization (MOO) framework. The optimized sentiment vector obtained at the end is used for the training of SVM for sentiment classification. The result analysis is performed on four Hindi datasets covering varying domains.

K. M. Anil Kumar et al [4] proposed a model for retrieving user’s sentiments from Kannada Web documents. Machine translation was used to translate the English reviews into Kannada, further POS tagger is used to implement adjective analysis and Turney algorithm which focuses on pair of POS. The polarity is considered as the difference between the positive and negative counts. If the value results more than zero then considered as positive, less than zero then negative else neutral.

Namita Mittal et al [5] developed an efficient approach based on negation and discourse relation for predicting sentiment. They improved HSWN by adding more opinion words to it. They proposed rules for handling negation and discourse that affected the prediction of sentiments. 80% accuracy was achieved by their proposed algorithm.

M. Farhadloo et. al. [6] proposed multiclass sentiment analysis for English language using clustering and score representation. The model used aspect level sentiment analysis. Bag of nouns was preferred instead of bag of words to enhance clustering results, score representation and more accurate sentiment identification.

Das and Bandopadhya[8] gave four strategies to predict sentiment of word. First strategy proposed by them was an interactive game which returned annotated words with their polarity. In second strategy, they use bi-lingual English and other Indian Language dictionaries to predict the polarity. In third approach, they use wordnet and synonym-antonym relation to predict the polarity. In fourth approach, polarity is determined by learning from pre-annotated corpora.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 2, February 2019

Joshi et al. [10] proposed fall back strategy for Hindi Language. Their strategy follows three approaches: In-Language Sentiment Analysis, Machine Translation, Resource based sentiment analysis. They developed Hindi SentiWordnet(HSWN) by replacing words of English SentiWordnet by their Hindi Equivalents. Final accuracy achieved by them is 78.14.

## III. METHODOLOGY

In this section, the proposed method is illustrated. First, generation on tourism review dataset is prepared which is further used for sentiment analysis. The proposed algorithm mainly consists of three phases i.e. Dataset Preparation, Feature Extraction and Classification which are discussed below:

### *A. Dataset Preparation*

In this subsection of research methodology, dataset is prepared. For data preparation first of all data is gathered/ acquisited from the online sources and then dataset is pre-processed to be useful for further proceedings. So, it contains basically :

Step 1: Data Gathering

Step 2: Text Preprocessing

#### *1) Data Gathering*

The very first step of proposed research methodology is dataset preparation in which data gathering is first of all performed. Figure 1 represents this step in which manually analysed reviews are collected and saved in a dataset format.

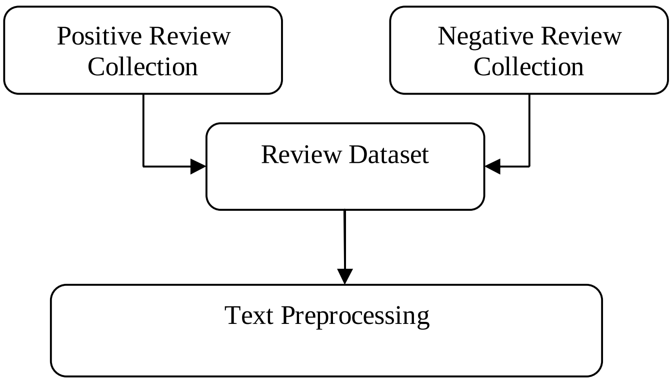


**Figure 1: Data Gatheringfor Proposed Work**

For data gathering is TripAdvisor website is used. From this website different reviews are collected from different tourist places. From this website 225 positive reviews and 225 negative reviews are manually collected and stored in separate database. These reviews are further processed to be identified by machine accurately.

#### *2) Text Pre-processing*

As it is necessary to clean the dataset. This process is termed as data preprocessing. This step is necessary to remove unnecessary terms used by an individuals in reviews such as comma, full-stop, colons or any special symbols or characters. These terms doesn't associate any sentiment values. So, to reduce further complexity it is needed to removes such terms. In this paper this step is performed in two steps as stated below:

Sometimes url is also mentioned in reviews as suggested by any reviewer. As it is known that any url represents address, they don't convey any emotional or sentimental values.

Some special symbols or characters are removed.

Comma, full stops, colons and semi-colons are removed.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

**International Journal of Innovative Research in Computer and Communication Engineering**

*(A High Impact Factor, Monthly, Peer Reviewed Journal)*

*Website: www.ijircce.com*

**Vol. 7, Issue 2, February 2019**

*B. Feature Extraction*

Extraction of some useful information out of these reviews is known as feature extraction. This information is a set of features that is very useful for classifiers. Figure 2 illustrates the process of feature extraction for proposed methodology. Basically, three features are extracted in this research that are discussed below.

*1) POS Occurrence Count*

In this sub-section, part of speech are differentiated for each lexicons in each reviews. For extracting POS of each lexicon Part-of-Speech Tagger (POS Tagger) is used. Each word in the sentence is termed as token and its POS is tagged i.e. whether it is noun, verb, pronoun, adverb, etc. In this paper Stanford POS tagger is referred and 48 POS tags are reduced to 11 POS tags which is shown in table 1.

**Table 1: Look-up Table for POS**

| Penn POS Tag | Description | Equivalent Proposed Algorithm Tag |
|---|---|---|
| JJ | Adjective | ADJ |
| JJR | Adjective, Comparative | ADJ |
| JJS | Adjective, Superlative | ADJ |
| NN | Noun, Singular | NN |
| NNS | Noun, Plural | NN |
| NNP | Proper Noun, Singular | NNP |
| NNPS | Proper Noun, Plural | NNP |
| RB | Adverb | ADV |
| RBR | Adverb, Comparative | ADV |
| RBS | Adverb, Superlative | ADV |
| VB | Verb | VB |
| VBD | Verb, past tense | VB |
| VBG | Verb, gerund or present participle | VB |
| VBN | Verb, past participle | VB |
| VBP | Verb, non-3$^{rd}$ person singular present | VB |
| VBZ | Verb, 3$^{rd}$ person singular present | VB |
| DT | Determiner/ Article | DT |
| IN | Preposition | PP |
| PR | Pronoun | PN |
| CC | Coordinating conjunction | CC |
| UH | Interjection | INJ |
| FW, MD, TO, PDT | Foreign word, Modal, to, Predeterminer | OTH |

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

**International Journal of Innovative Research in Computer and Communication Engineering**

*(A High Impact Factor, Monthly, Peer Reviewed Journal)*

**Website: www.ijircce.com**

**Vol. 7, Issue 2, February 2019**

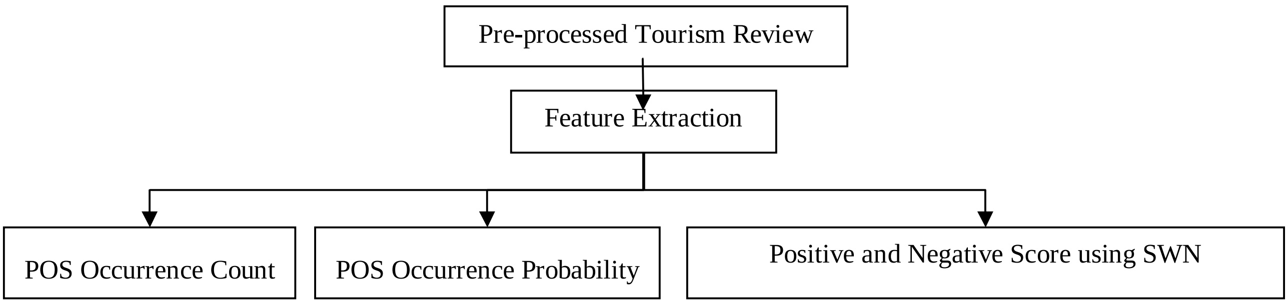


**Figure 2: Feature Extraction for Proposed Work**

To illustrate the working of this step an example is taken as: "The room is maintained very well and clean". After using Penn POS Tags, the given example has following POS tagging:

| | |
|---|---|
| The | 'DT' |
| Room | 'NN' |
| Is | 'VB' |
| Maintained | 'VB' |
| Very | 'ADV' |
| Well | 'ADV' |
| And | 'CC' |
| clean | 'VB' |

POS Occurrence Count is then found by counting the time of occurrence of each part of speech in the sentence. For above example POS occurrence count is determined as below. So. In this way 1st feature vector is prepared.

| | |
|---|---|
| Verb | 3 |
| Other | 0 |
| Determiner | 1 |
| Noun | 1 |
| Conjunction | 1 |
| Adjective | 0 |
| Proper Noun | 0 |
| Adverb | 2 |
| Pronoun | 0 |
| Interjection | 0 |
| Preposition | 0 |

*2) POS Occurrence Probability*

After POS occurrence count, POS occurrence probability is calculated. This probability is determined by occurrence of each POS out of total words is found and can be calculates as:

$$\text{Probablity\_POS} = \frac{\text{Count of POS}}{\text{Total Number of words in review sentence}} \qquad \text{(i)}$$

For above example POS occurrence probability is as follows:

| | |
|---|---|
| Verb | 3/8 = 0.375 |
| Other | 0/8 =0 |
| Determiner | 1/8 =0.125 |
| Noun | 1/8 =0.125 |
| Conjunction | 1/8 =0.125 |
| Adjective | 0/8 =0 |
| Proper Noun | 0/8 =0 |
| Adverb | 2/8 = 0.25 |

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 2, February 2019

| | |
|---|---|
| Pronoun | 0/8 =0 |
| Interjection | 0/8 =0 |
| Preposition | 0/8 =0 |

*3) Positive and Negative Score Feature using SWN*

For generation of feature vector associated with positive and negative score of the review sentence. Sentiword Netdictionary is referred. In this paper sentiment score is calculated by summarizing the positive and negative score of each word in entire sentence. If the positive score of entire sentence is greater than negative score of entire sentence then the overall sentiment score of that sentence/review is positive. Similarly if the positive score of entire sentence is less than negative score of entire sentence then the overall sentiment score of that sentence/review is negative. The calculation of positive and negative score of the sentence is determined as follows:

$$\text{Positive\_Score\_Sentence} = \frac{\text{Sum of Positive Score of Each Word}}{\text{Total Number of words in review sentence}} \quad \text{(ii)}$$

$$\text{Negative\_Score\_Sentence} = \frac{\text{Sum of Negative Score of Each Word}}{\text{Total Number of words in review sentence}} \quad \text{(iii)}$$

For Generation of Positive and Negative Score using SWN in above example is as follows:

| Word | Pos Score | Neg Score |
|---|---|---|
| The | 0 | 0 |
| Room | 0.016243 | 0.016243 |
| Is | 0.043266 | 0.071605 |
| Maintained | 0 | 0 |
| Very | 0.080357 | 0.044643 |
| Well | 0.195412 | 0.051914 |
| And | 0.029759 | 0.030189 |
| clean | 0.107692 | 0.115385 |

So, the overall positive score of the sentence is 0.472 and negative score is 0.329 which represents that this sentence is of positive type So, all the three feature vectors are combined and form a feature vector having 24 features and 1 class label.

*C. Classification*

Classification technique is used to categorize any data values into different classes. In this paper deep learning approach is used to classify dataset into positive or negative class. Recently, in NLP, deep learning is an area of interest. In deep neural network, entire sentence as feed as input in network and each word is processed and finally output is observed.

In this research methodology long short term memory (LSTM) deep network is used for classification. After feature vector creation the dataset is divided into training and testing ratio in 70:30 ratio and feed into LSTM. The training dataset is provided with their class labels and trained network is created over which further testing is performed to predict the final review tag as positive or negative.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 2, February 2019

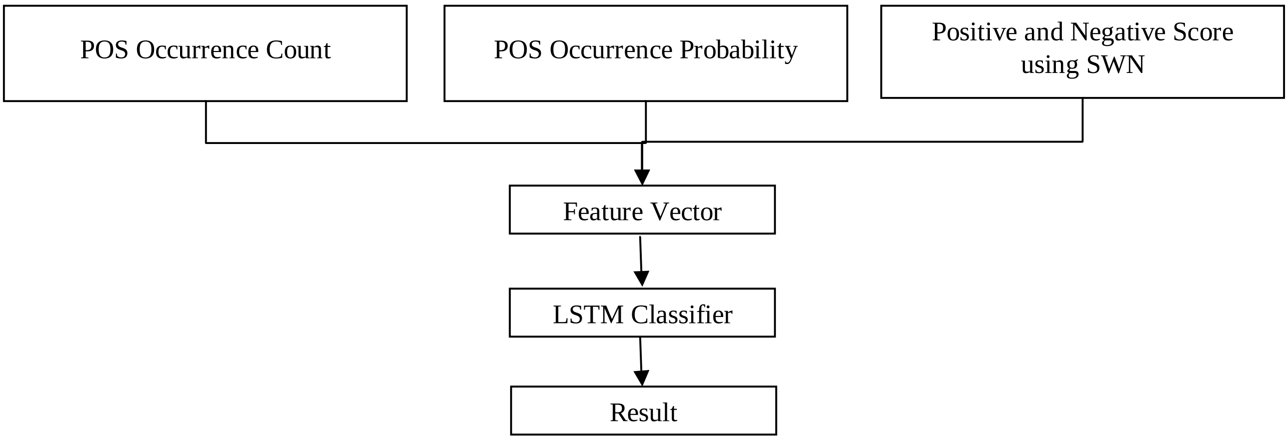


**Figure 3: Feature Classification for Proposed Work**

*1) Long Short Term Memory (LSTM)*

One of the type of recurrent neural network (RNN) is Long short-term memory (LSTM). It is type of deep learning neural network approach composed of several neural network modules. The LSTM network is composed of four units: memory cell, input unit, output unit and forget unit. The memory unit is the unit that stores the data values for some time intervals and remaining three units regulates the flow of data values for evaluation of output value. The LSTM deep network is used for both classification as well as regression process.

At each time step t there is a set of vectors, including an input gate it, a forget gate $f_t$, an output gate $o_t$ and a memory cell $C_t$.

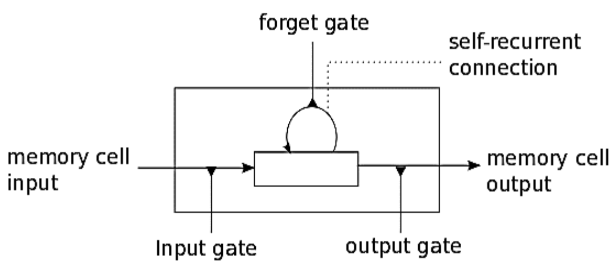


**Figure 4: Long Short-Term Memory (LSTM) Units**

All these together are used to compute the output of the hidden layer ht as follows:

$$f_t = \sigma(W_f * x_t + U_f * h_{t-1} + b_f) \quad \text{(iv)}$$

$$i_t = \sigma(W_i * x_t + U_i * h_{t-1} + b_i) \quad \text{(v)}$$

$$\hat{C}_t = tanh(W_c * x_t + U_c * h_{t-1} + b_c) \quad \text{(vi)}$$

$$C_t = i_t * \hat{C}_t + f_t * C_{t-1}) \quad \text{(vii)}$$


DOI: 10.15680/IJIRCCE.2019. 0702019

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 2, February 2019

$$o_t = \sigma(W_o * x_t + U_o * h_{t-1} + b_o) \quad \text{(viii)}$$

$$h_t = o_t * \tanh(C_t)) \quad \text{(ix)}$$

In this model, σ is the sigmoid activation function, tanh the hyperbolic tangent activation function, $x_t$ the input at time t, $W_i$ , $W_C$, $W_f$ , $W_o$, $U_i$ , $U_C$ , $U_f$ , $U_o$ are weight matrices to regulate the input and $b_i$ , $b_C$ , $b_f$ , $b_o$ are bias vectors.
Four steps of the LSTM network are discussed below:

Step A
First, the model needs to determine what to throw away from the cell State. This is referred to as the forget gate values ft. The input in this step is the output of the previous step ht−1 and the input xt. A sigmoid activation function is used to give output values between 0 and 1, where 0 corresponds to "let nothing through" and 1 to "remembering everything".

Step B
The next step is to determine what information is going to be added to cell state. In this step again the inputs are $h_{t-1}$ and $x_t$. The input layer gate it first applies a sigmoid layer over the input to determine which parts of the cell state will be updated. Then a tanh layer is used to create new candidate values $C^{\sim}_t$. In the next step, these two will be combined to update the cell state $C_t$.

Step C
Now the old cell state is multiplied by ft, to forget the things that are not needed anymore and the new information is added to the cell state memory.

Step D
In the final step, it is determined what the output $h_t$ is. First, a sigmoid layer is applied to the previous output $h_{t-1}$ and input $x_t$, to determine the output gate values $o_t$. This is a value between 0 and 1 indicating which parts of the cell state are going to be output. Then the cell state $C_t$ is transformed by a tanh function to get values between -1 and 1.
Overall proposed algorithm is shown in figure 5 which include three phases i.e. Dataset Preparation, Feature Extraction and Classification.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 2, February 2019

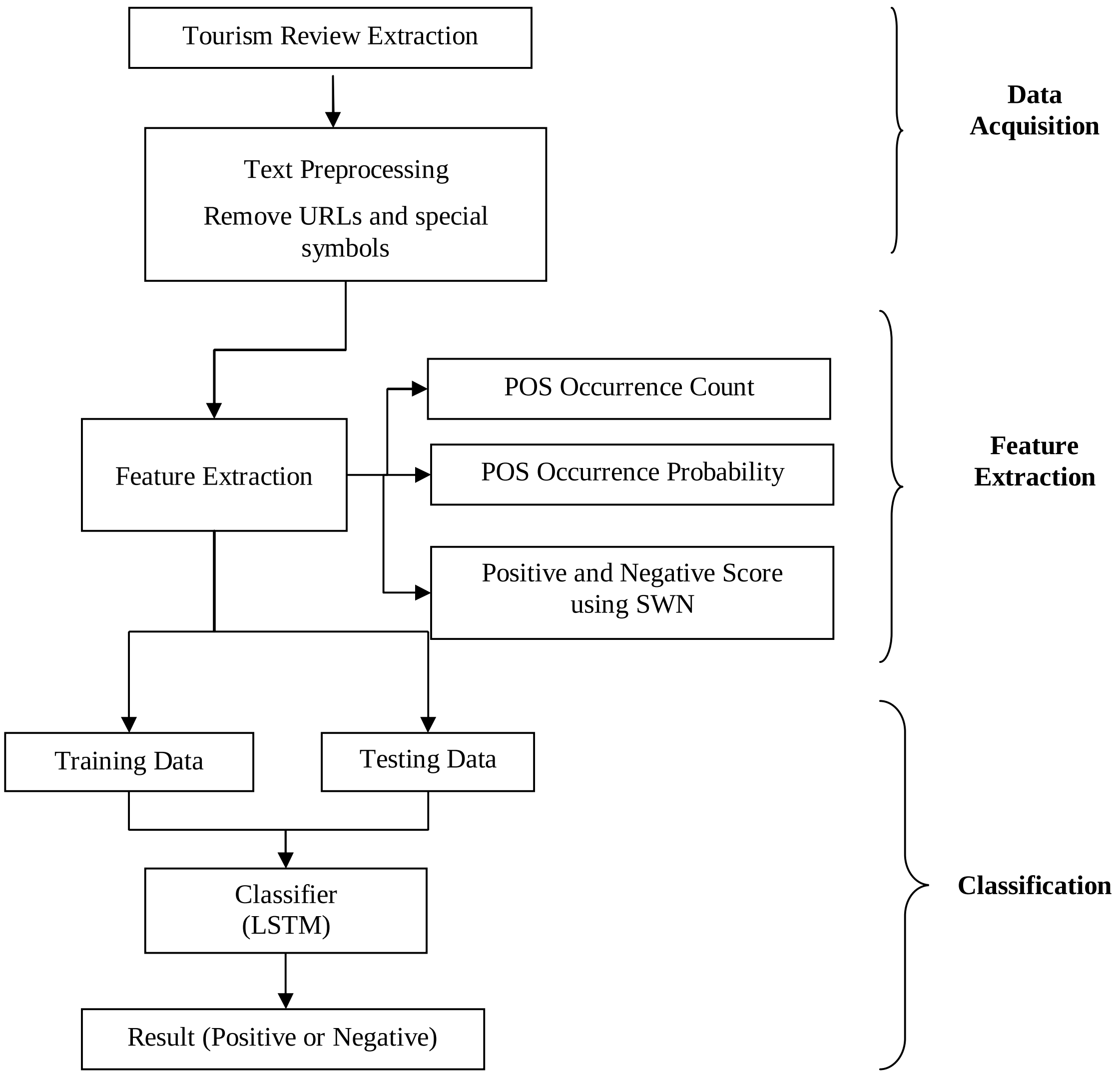


**Figure 5: Proposed Flow Diagram of Tourism Review Sentiment Analysis**

## IV. RESULT ANALYSIS

For simulation result, the research is focused towards lexicon feature extraction for sentiment analysis from reviews. For executing this simulation, tourism review dataset is prepared with 450 reviews given on tripadvisor website. For this the research methodology is performed in four steps. In the first step lexicons are extracted from the review and their Part-Of-Speech (POS) are tagged. In the second step the extracted part of speech tags occurrence are extracted. In third step, each POS tag's sentiment score calculated by referring to the SentiWordNet (SWN) dictionary. Subsequently, in the last step, the extracted features are fed into LSTM deep learning classifier to classify the review into either positive and negative type of review. After classification four performance measures are evaluated. These evaluations are performed on testing dataset. The training testing dataset is divided into 70:30 ratio. Some of the performance parameters are discussed below:
Accuracy is represented as:

$$(TP+TN)/(TP+TN+FP+FN) \qquad (x)$$

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 2, February 2019

Precision is represented as:

$$(TP)/(TP+FP) \quad \text{(xi)}$$

Recall is represented as:

$$(TP)/(TP+FN) \quad \text{(xii)}$$

F_measure is represented as:

$$(2*Recall*Precision)/(Recall + Precision) \quad \text{(xiii)}$$

Where,
TP= True Positive, that means if one sample is anpositive and the predicted label also stands positive.
TN= True Negative, that means if one sample is negative and the predicted label stands negative.
FP = False Positive, that means if one sample is negative and the predicted label stands positive.
FN= False Negative, that means if one sample is positive and the predicted label stands negative.
TP stands the number of true positive samples, FN stands the number of false negative samples, FP stands the number of false positive samples, and TN stands the number of true negatives.
Performance evaluation of proposed research methodology is evaluated in table 2. As proposed methodology is designed for sentiment analysis using textual features from reviews. After extraction of features from reviews machine learning approaches such as support vector machine (SVM), k-nearest neighbor(k-NN) and long short term memory (LSTM) classifiers to classify into positive and negative reviews. After simulation results it is seen that LSTM have highest accuracy, recall, precision and f_measure as compared to SVM and k-NN.

**Table 2: Performance Evaluation of Proposed Algorithm**

| | KNN | SVM | LSTM |
|---|---|---|---|
| Accuracy | 83.7037 | 89.6296 | 98.5185 |
| Recall | 76.5625 | 81.25 | 96.875 |
| Precision | 87.5 | 96.2963 | 100 |
| F_Measure | 81.6667 | 88.1356 | 98.4127 |

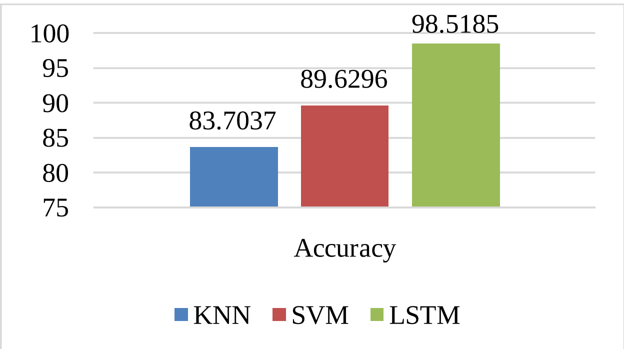


**Figure 6: Performance Comparison of Accuracy**

Figure 6 represents performance comparison of LSTM , support vector machine and k-NN classifier for sentiment classification. After result analysis it is observed that LSTM gives better accuracy as compared to other classifiers.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

**International Journal of Innovative Research in Computer and Communication Engineering**

*(A High Impact Factor, Monthly, Peer Reviewed Journal)*

Website: www.ijircce.com

Vol. 7, Issue 2, February 2019

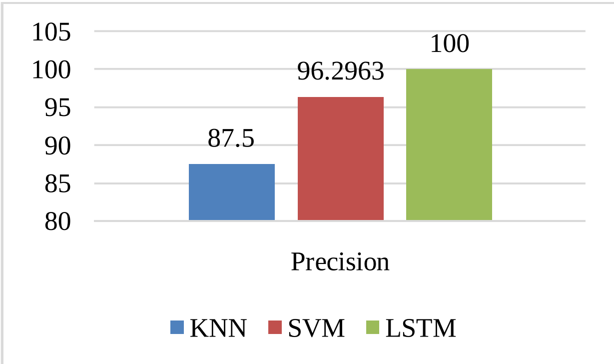


**Figure 7: Performance Comparison of Precision**

Figure 7 represents performance comparison of LSTM, support vector machine and k-NN classifier for sentiment classification. After result analysis it is observed that LSTM gives better precision as compared to other classifiers.

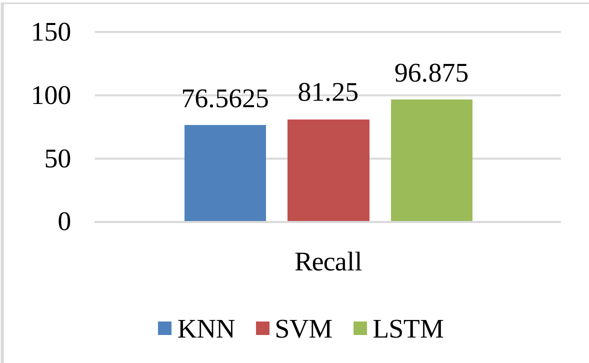


**Figure 8: Performance Comparison of Recall**

Figure 8 represents performance comparison of LSTM, support vector machine and k-NN classifier for sentiment classification. After result analysis it is observed that LSTM gives better recall as compared to other classifiers.

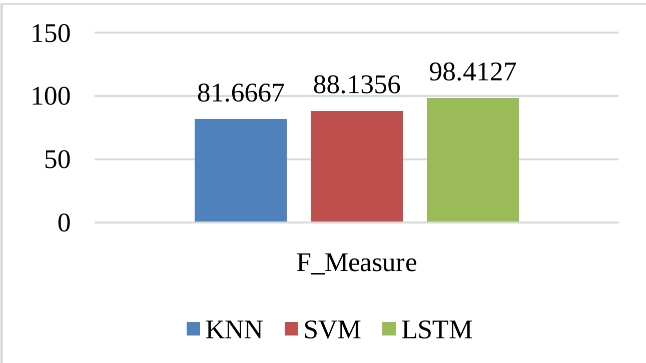


**Figure 9: Performance Comparison of F_Measure**

Figure 9 represents performance comparison of LSTM, support vector machine and k-NN classifier for sentiment classification. After result analysis it is observed that LSTM gives better f_measure as compared to other classifiers.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 2, February 2019

**Table 3: Comparative Performance Evaluation of Proposed Algorithm**

| | LSTM | Sruthi S [1] |
|---|---|---|
| **Accuracy** | 98.5185 | 95.43 |
| **Recall** | 96.875 | 95 |
| **Precision** | 100 | 96 |

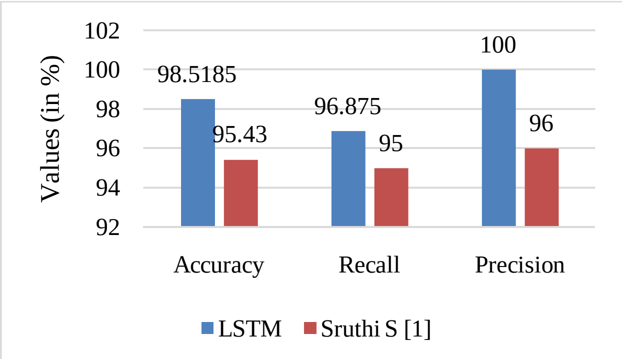


**Figure 10: Comparative Performance Analysis with Existing Work**

Table 3 and figure 10 represents the comparison for binary classification of existing techniques as well as proposed techniques. The binary classification is for positive data class as well as negative data class. Existing Work is performed using lexicon features only and Naïve bayes classifier performance are evaluated in existing work. Whereas in proposed work lexicon as well as sentiscore are considered and probability of occurrence of POS is evaluated and classified using LSTM classifier. It has been observed that proposed algorithm outperforms better as compared to existing work

## V. CONCLUSION

In this paper, a domain-specific lexicon generation method is proposed for tourism review sentiment analysis. A completed lexicon-based sentiment classification framework is proposed.In this paper, a sentiment analysis framework is proposed for tourist review to classify reviews either in positive sense or negative sense. For this the research methodology is performed in four steps. In the first step lexicons are extracted from the review and their Part-Of-Speech (POS) are tagged. In the second step the extracted part of speech tags occurrence are extracted. In third step, each POS tag's sentiment score calculated by referring to the SentiWordNet (SWN) dictionary. Subsequently, in the last step, the extracted features are fed into LSTM deep learning classifier to classify the review into either positive and negative type of review.For the result analysis, the dataset is prepared by collecting reviews from tripadvisor website and performance is analyzed using classifiers such as support vector machine (SVM), k nearest neighbor (KNN) and LSTM deep network. The simulation gives output (accuracy) for SVM about 84%, k-NN about 90% and LSTM about 98% respectively.In this paper a comparison with existing work is also performed using Naïve bayes classifier and it has been observed that proposed algorithm outperforms better as compared to existing work.Future work can be pursued in several directions. The proposed sentiment analysis system can be integrated with multiclassification with deep learning that can exploit advantages of current models and address their limitations.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 2, February 2019

## REFERENCES

[1] Sruthi S, Reshma Sheik and Ansamma John, “Reduced Feature Based Sentiment Analysis on Movie Reviews Using Key Terms”, IEEE, 2017.
[2] Tirath Prasad Sahu and Sanjeev Ahuja, “Sentiment Analysis of Movie Reviews: A study on Feature Selection & Classification Algorithms”, IEEE, 2016.
[3] Md Shad Akhtar, Ayush Kumar, Asif Ekbal, Pushpak Bhattacharyya, “A Hybrid Deep Learning Architecture for Sentiment Analysis”, International Conference on Computational Linguistics: Technical Papers, pp. 482–493, 2016.
[4] K. M. Anil Kumar, N. Rajasimha, M Reddy, A. Rajanarayana, K. Nadgir, “Analysis of Users’ Sentiments from Kannada Web Documents”, International Conference on Communication Networks, vol. 54, pp. 247-256, 2015.
[5] Namita Mittal, Basant Aggarwal, Garvit Chouhan, Nitin Bania, Prateek Pareek, “Sentiment Analysis of Hindi Review based on based on Negation and Discourse Relation”, International Joint Conference on Natural Language Processing, pp 45-50, 2013.
[6] Mohsen Farhadloo, Erik Rolland,” Multi-Class Sentiment Analysis with Clustering and Score Representation”, IEEE 13th International Conference on Data Mining Workshops, pp. 904-912, 2013.
[7] Singh, V. K., et al. "Sentiment analysis of movie reviews: A new feature based heuristic for aspect-level sentiment classification." Automation, Computing, Communication, Control and Compressed Sensing (iMac4s), 2013 International Multi-Conference on. IEEE, 2013.
[8] Amitava Das, Sivaji Bandopadaya, “SentiWordnet for Bangla”, Knowledge Sharing Event -4: Task, Volume 2,2010.
[9] Amitava Das, Sivaji Bandopadaya, “SentiWordnet for indian language”, Workshop on Asian Language Resources, pp. 56-63, Beijing, China, 21-22 August 2010.
[10] Aditya Joshi, Balamurali AR, Pushpak Bhattacharya, “A fall back strategy for sentiment analysis in hindi”, International Conference on Natural Language Processing, 2010.
[11] Annett, Michelle, and GrzegorzKondrak. "A comparison of sentiment analysis techniques: Polarizing movie blogs." Advances in artificial intelligence. Springer Berlin Heidelberg, 2008. 25-35.