

Aspect based Nepali Text Sentiment Analysis using BiLSTM

Piyush Pant¹, Subarna Shakya²

Department of Electronics and Computer Engineering, Institute of Engineering, Tribhuvan University, Lalitpur, Nepal

E-mail: ¹075mcsk011.piyush@pcampus.edu.np, ²drss@ioe.edu.np

Abstract

Aspect-Based Sentiment Analysis is a text analysis technique that classifies the sentiment of a specific aspect in a text. With a rise in internet and social media access, people tend to decide to buy a product or to visit a restaurant based on comments and reviews. Categorical reviews can play a significant role in improving the business of a restaurant if a business entity can capture insights of those reviews and use productively. Since Nepali is a low resource language there is no significant numbers of the dataset for restaurant reviews. Therefore a dataset is created in the work to model a target aspect-based sentiment analysis in restaurant review domain. The dataset comprises of comments extracted from social media sites, websites, and manually prepared data. Data are cleaned, pre-processed then aspects and sentiments are extracted using POS tagging. Sentiments and target categories are classified using word embedding and the deep learning BiLSTM model. The sentiment classification model classifies the sentiments into positive or negative. The target classification model classifies text into 5 target categories Food (F), Price (P), Service (S), Hygiene (H), and Experience (E). The result was found to be good compared to similar models in other domains. The sentiment prediction model and target category prediction model attains the F1 Score of 0.88 and 0.76 respectively.

Keywords: ABSA, BiLSTM, LSTM, POS, TABSA

1. Introduction

Sentiment analysis aims to extract sentiments and opinions from any natural language text through computational methods. It impacts the lives of people by providing technology to collect customer feedback on products and services which are used to improve the quality of products and services. Applications of sentiment analysis on

different domains such as e-commerce, business intelligence, political analysis, market research, product analysis, and social media monitoring have made it one of the hottest topics to research on. Document level sentiment analysis emphasize a single sentiment for the entire document; The sentence level analysis is a different kind of analysis that provides a sentiment for each sentences, and the aspect level provides sentiments of entities or phrases within the sentence. Aspect Based Sentiment Analysis (ABSA) has opened doors to identify the various aspects based on a given text. Target Based Sentiment Analysis (TABSA) is a variant of ABSA which is used to detect different aspects based on the given target. Aspect Based Sentiment Analysis solves the problem of sentiment analysis of not being able to capture fine detail data towards specific aspects and has attracted many researchers. Research on aspect extraction and sentiment analysis according to aspect terms has significant applications in the real world. Nepali is morphologically rich, under-resourced, and a free word order language. There are only a very few tools and annotated corpus available for the Nepali language. There is notable research in the Nepali language to identify the aspects and sentiments from texts. The Nepali language is similar to the Hindi language, but they differ in morphological rules, vocabulary, semantics, and grammar rules. Online reviews and comments make a significant impact on people perception to buy a product or visiting a restaurant. Nowadays, people decide to buy products or plan a trip based on reviews. Restaurant businesses can benefit by the analysis of these reviews, and they can improve their strategy to have more customers and profit.

1.1 Problem Statement

There have been number of researches in sentiment analysis and aspect-based sentiment analysis worldwide. Sentiment Analysis in Nepali language is still in development phase. Analysing sentiments of Nepali comments given in e commerce platforms can significantly help e commerce to grow business. Restaurant can benefit by feedback of aspect based sentiment analysis of restaurant review data. We are trying to design a dataset and model which can do aspect-based sentiment analysis on more fine grained aspects of Nepali comments using BiLSTM in e commerce platform, focusing on food domain and restaurant review.

2. Related Work

The wide application domains of sentiment analysis have attracted many researchers. Earlier research was primarily focused on word level and sentence level analysis of a text.

This could not analyse deep-level semantics and fine-grained details. Aspect Based Sentiment Analysis (ABSA) evolved as the next research area to overcome these problems. ABSA can analyse based on the sentiment of text on predefined categories of the text. Aspect-based sentiment analysis is an advanced text analysis technique that segregates text into aspects and then allocates each one a sentiment level (positive or negative). Generally, aspects are nouns, noun groups, or proper nouns whereas sentiments are adjectives. There is research on aspect term sentiment analysis in languages like Hindi. [1] Utilized Support Vector Machine (SVM) and Conditional Random Field (CRF) for a sentiment classification process with the aspect term extraction on the Hindi language. There is research in sentiment analysis of Nepali texts. [2] Developed Nepali SentiWordNet Corpus developed based on the subjective and objective level of the sentences with a binary value annotation. It also structured with Bhavanaks (Nepali SentiWordNet) model which is a translated model of English WordNet. [3] Analysed Nepali text using machine learning classifiers SVM and Naive Bayes. The Nepali text were collected from different websites and with a POS tagging for extracting the sentiment and aspect terms. Initially the data were labelled through manual process, and they used SVM and Naive Bayes to classify the sentence. [4] Used multilingual BERT model along with the BiLSTM for extracting the aspect term for sentiment classification process and achieved an F1 scores of 0.5798 and 0.8150 respectively. LSTM preserves information from inputs that have passed through it using a hidden state. But it only preserves information from the past. BiLSTM can preserve information from both the past and future which provides greater accuracy in learning information.

3. Proposed Work

3.1 DataSet

Table 1 shows the sample of manual data prepared and tagged with category and polarity. Few samples from dataset are shown in Table 1.

Dataset consists of 4000 rows of filtered and tagged data. The polarity distribution of data is as shown in Figure 1. The count of the number of rows of positive and negative comments are 2345 and 1755 respectively.

The category distribution of data is as shown in Figure 2. The count of number of rows of Experience(E), Service(S), Hygiene(H), Food(F) and Price(P) is 945, 795, 452, 1360, and 448 respectively.

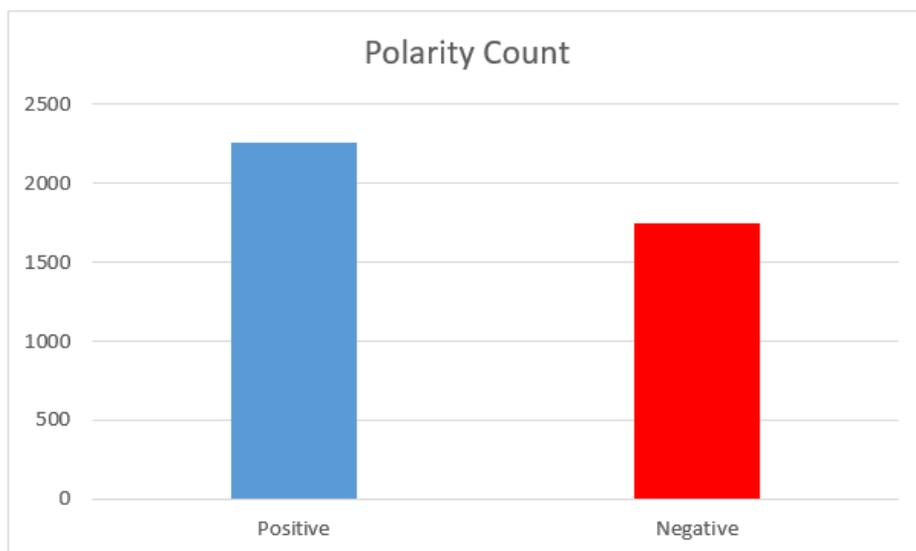


Figure 1. Polarity Distribution of Dataset

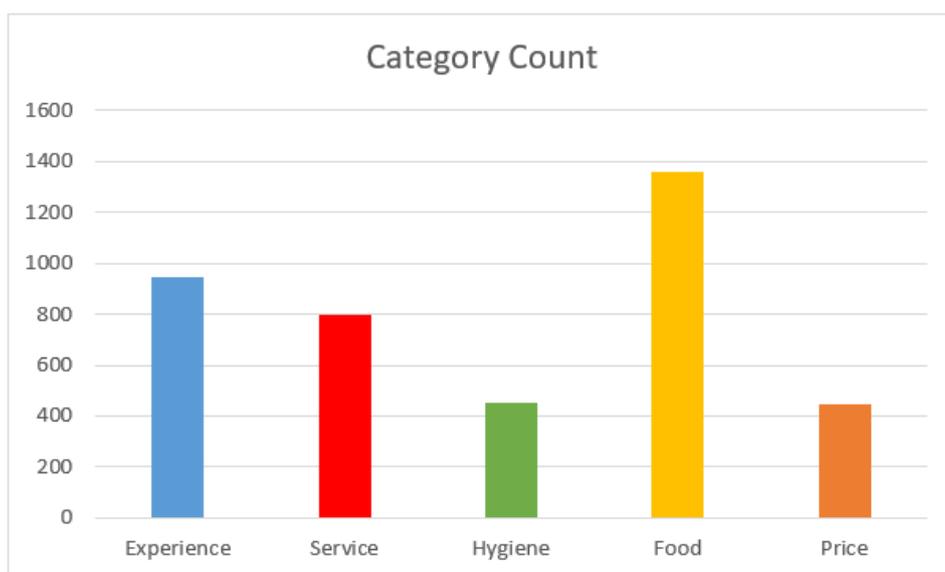


Figure 2. Category Distribution of Dataset

Table 1. Few Samples from Dataset

Review	Category	Polarity
खाना सस्तो थियो	P	0
शान्त रेस्टुरेन्ट	E	1

निराशाजनक अनुभव।	E	0
तरकारी नमिठो थियो	F	0
खाना धेरै उत्कृष्ट थियो	F	1
अनुशासित कर्मचारी	S	1
कस्तो फोहोर शौचालय	H	0

3.2 Architecture Design

The architecture design is shown in Figure 3.

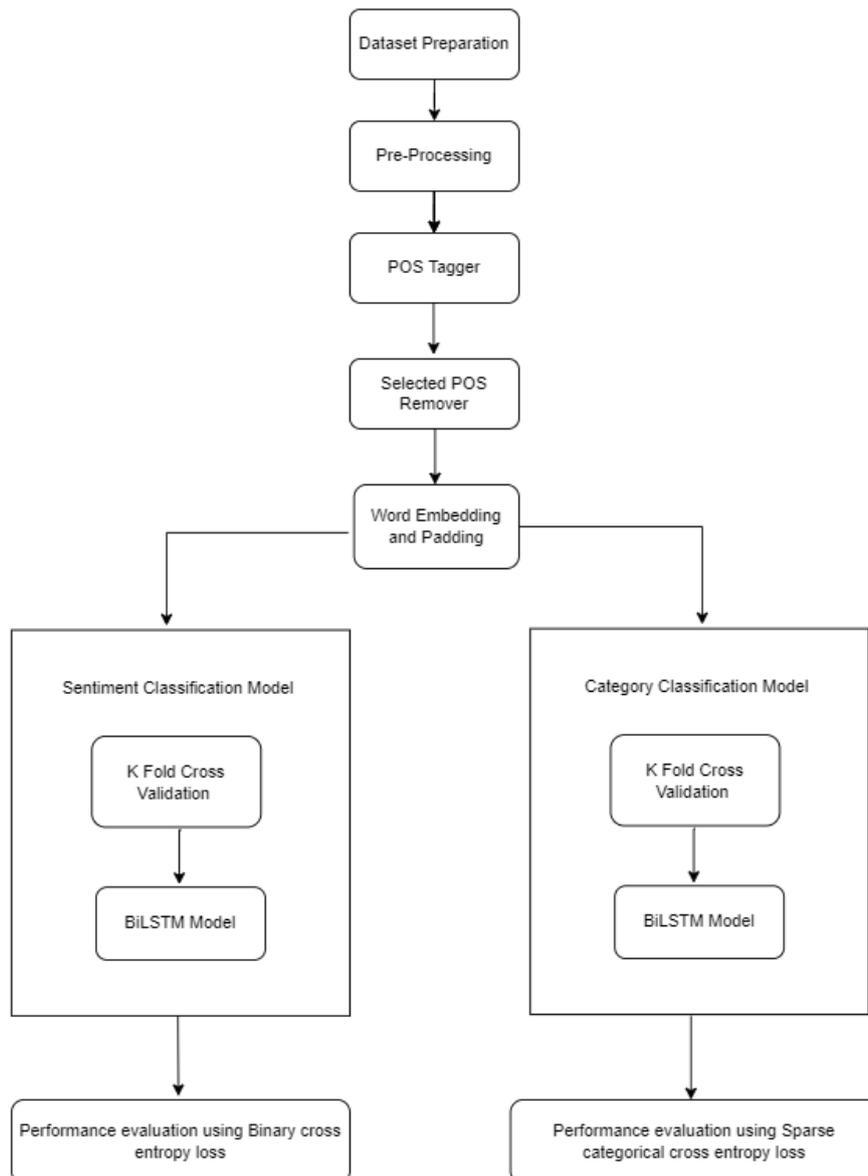


Figure 3. System Architecture Design for Aspect based Nepali Text Sentiment Analysis using BiLSTM

For the preparation of dataset, we first started web scraping data from multiple web pages and social sites. Food-related data were scrapped from Facebook pages such as foodmandu and Nepal.food. An Instant Data scrapper was used to extract the comments. Most of the comments extracted contain romanized texts. Language detection of the comments was done with langdetect python library and then Nepali language text was filtered. Among 3000 rows extracted only 300 rows of data were Nepali text. We extracted 300 rows of Nepali text following this process. For the rest of the data preparation, we utilized comments of 'www.tripadvisor.com'. We converted English comments of 'www.tripadvisor.com' to Nepali text using googletrans library. We corrected the grammar mismatch of the converted comments and extracted 1500 data using this process. The rest of the 2200 data were manually prepared. Then manual tagging of two polarity terms Positive and Negative (1 and 0) and 5 category terms Food(F), Experience(E), Service(S), Price(P), and Hygiene(H) was performed.

Data Pre-processing and cleaning is done to remove non relevant information from our dataset. Unwanted information from data which doesn't corresponds to meaning are removed by removing URL's, Digits, Tags and Mentions. Stop words which affect the meaning of a sentence like छ, छैन, थिएन, राम्रो, सही are kept and all the others are removed from text. Words are lemmatized to get the root word and reduce to inflectional forms. We used Nepali-nlp3 library for lemmatizing words.

POS tagger is prepared by modifying the nepali.pos data corpus in NLTK Corpus [5]. A restaurant review related dataset is added in nepali.pos tagger. The POS tagger tags dataset into different tags like DUM, NN, JJ, JJM, POP, VBO, VBKO, VBI, VBX, VBNE, and Unk. Aspects are indicated by noun words and sentiments are identified by adjective words. Auxiliary Verb (VBX), Normal Unmarked Adjective (JJ), Marked Adjective (J JM), Degree Adjective (J JD), Noun (NN) are filtered and kept and other POS tags are removed to prepare POS removed review. Auxiliary Verb (VBX) term makes positive or negative impact on meaning. Thus, these words are kept. Example of auxiliary verbs are follows: थियो, छ, भन्नुभयो, हो, छौं, छैन, थिए, होइन् and छन्. Figure 4 shows the example output of POS tagger.

Word embedding is a learned representation of text where words that have the same meaning have a similar representation. Word Embedding marks the words as vectors with real numbers. It is extracted from a large corpus and captures semantic knowledge from the corpus. NPVec1 [6] is the publicly available word embedding model prepared by training using 279-million-word tokens. It is the largest embedding designed for Nepali language. The word embedding model is the word2vec model which embeds the word into 300-dimension space. This model is used to convert our input data to vector space. The POS removed text is tokenized and padded with 60 words for each text. Embedding matrix is a matrix of all words and their corresponding embedding. We use an embedding matrix in an embedding layer in our model to embed a token into its vector representation, which contains information regarding that token or word.

Polarity	Review	Category	CleanedReviews	Tagged_words	Tagged_Noun	Tagged_Adj	Tagged_Unk	PosRemoved
1	खाना सस्तो थियो	P	खाना सस्तो	[[खाना, NN], (सस्तो, JJM)]	खाना	सस्तो		खाना सस्तो
1	शान्त रेस्टुरेन्ट	E	शान्त रेस्टुरेन्ट	[[शान्त, JJ], (रेस्टुरेन्ट, NN)]	रेस्टुरेन्ट	शान्त		शान्त रेस्टुरेन्ट
0	निराशाजनक अनुभव।	E	निराशाजनक अनुभव	[[निराशाजनक, JJ], (अनुभव, NN)]	अनुभव	निराशाजनक		निराशाजनक अनुभव
0	तरकारी नमिठो थियो	F	तरकारी नमिठो	[[तरकारी, NN], (नमिठो, ADJ)]	तरकारी	नमिठो		तरकारी नमिठो
1	खाना धेरै उत्कृष्ट थियो	F	खाना उत्कृष्ट	[[खाना, NN], (उत्कृष्ट, JJ)]	खाना	उत्कृष्ट		खाना उत्कृष्ट
0	अनुशासित कर्मचारी	S	अनुशासित कर्मचारी	[[अनुशासित, ADJ], (कर्मचारी, NN)]	कर्मचारी	अनुशासित		अनुशासित कर्मचारी
0	कस्तो फोहोर शौचालय	H	कस्तो फोहोर शौचालय	[[कस्तो, QW], (फोहोर, JJ), (शौचालय, NN)]	शौचालय	फोहोर		फोहोर शौचालय
1	उत्तम सेवा	S	उत्तम सेवा	[[उत्तम, JJ], (सेवा, NN)]	सेवा	उत्तम		उत्तम सेवा
1	कामदारहरु सफा छन्	H	कामदार सफा	[[कामदार, NN], (सफा, ADJ)]	कामदार	सफा		कामदार सफा
0	चाउमिन महेँगो थियो	P	चाउमिन महेँगो	[[चाउमिन, NN], (महेँगो, JJM)]	चाउमिन	महेँगो		चाउमिन महेँगो

Figure 4. POS Tagged Data

K Fold cross-validation is used to divide the dataset into 5 folds. In Each fold training and testing are done in 80:20 ratio. Model is prepared by passing embedding matrix to BiLSTM layer and Dense layer and the sigmoid activation function to transform output between 0 to 1. Another model is created for the target classification task in which Cleaned Reviews are trained with the 5 Categories. Categories are converted to integers and mapped. The model is prepared by using embedding matrix, BiLSTM model, Dense layer, and softmax function, and the model is evaluated. Validations on test data is done using this model as well.

For measuring the performance of the polarity and category classification model we used various evaluation matrices like precision, accuracy, f1-scores, and recall and loss functions. Binary cross entropy loss is used to evaluate the performance of sentiment model which predicts output as positive and negative. Sparse categorical cross entropy is used to

evaluate the performance of category classification model which classifies output into 5 classes Food (F), Price (P), Service (S), Hygiene (H), and Experience (E).

4. Results and Discussion

4.1 Results

The experiments were carried out on jupyter notebook. The dataset specified in Table 1 was applied to the system architecture provided in Figure 3. The parameters of the models prepared are as shown in Table 2.

Table 2. Parameters of Model

Parameters	Sentiment Model	Category Model
Batch Size	128	128
Epoch	10	10
Hidden Layer Optimizer	Relu	relu
Output Layer Optimizer	Adam	adam
Activation Function	Sigmoid	softmax
Loss Function	binary cross entropy	sparse categorical cross entropy

Table 3. Cleaned and Aspect Extracted Data

Review	Cleaned Review	Aspect	Sentiment
खाना सस्तो थियो	खाना सस्तो	खाना	सस्तो
शान्त रेस्टुरेन्ट	शान्त रेस्टुरेन्ट	रेस्टुरेन्ट	शान्त
निराशाजनक अनुभव।	निराशाजनक अनुभव	खाना	स्वस्थ
तरकारी नमिठो थियो	तरकारी नमिठो	ठाउँ	उत्तम
खाना धेरै उत्कृष्ट थियो	खाना उत्कृष्ट	मूल्य	उचित

अनुशासित कर्मचारी	अनुशासित कर्मचारी	कर्मचारी	अनुशासित
कस्तो फोहोर शौचालय	कस्तो फोहोर शौचालय	शौचालय	फोहोर

Firstly, reviews in the dataset Table 1 are pre-processed and cleaned to get the cleaned review. The idea behind removing stop words was to remove the unwanted words which don't add meaning to the text and keep only meaningful information. Then, POS tagging is done to extract the POS tags of noun and adjective. VBX, JJ, JJM, JJD, NN are filtered and kept and other POS tags are removed to prepare POS removed review. Vocabulary is created from a POS removed reviews using the word2vec model NPVec1 [6]. Data is tokenized and padded to fix a length of 60 to make a consistent matrix. The embedding matrix is prepared from the input POS removed review. Nouns are chosen for aspect terms and adjectives are the sentiments of aspects. Table 3 shows sample data after cleaning, POS tagging, and Aspect term extraction.

The sentiment prediction model is trained with POS removed review with respect to polarity. KFold Cross-validation is used to evaluate the best sentiment prediction model using K=5. In Each Fold, data is split into training and testing sets with an 80:20 ratio. The dropout layer was added to remove the over-fitting in our model. Models were tested with different combinations of hidden layers, dropouts, epochs, and batch size. The final chosen combination results in the following output of Table 4.

Table 4. K Fold Cross Validation Result of Sentiment Prediction Model

Fold	Loss	Accuracy	Validation Loss	Validation Accuracy
1	0.2652	0.8946	0.3171	0.8708
2	0.2584	0.8938	0.3246	0.8644
3	0.2756	0.8858	0.3263	0.8676
4	0.2785	0.8830	0.3226	0.8644
5	0.2638	0.8934	0.3205	0.8676

Table 4 The first fold gave the best result having a minimum validation loss of 0.3171 and validation accuracy of 0.8708. This model was chosen for the sentiment classification task. The accuracy and loss in each fold are good which is showing the good performance of our model.

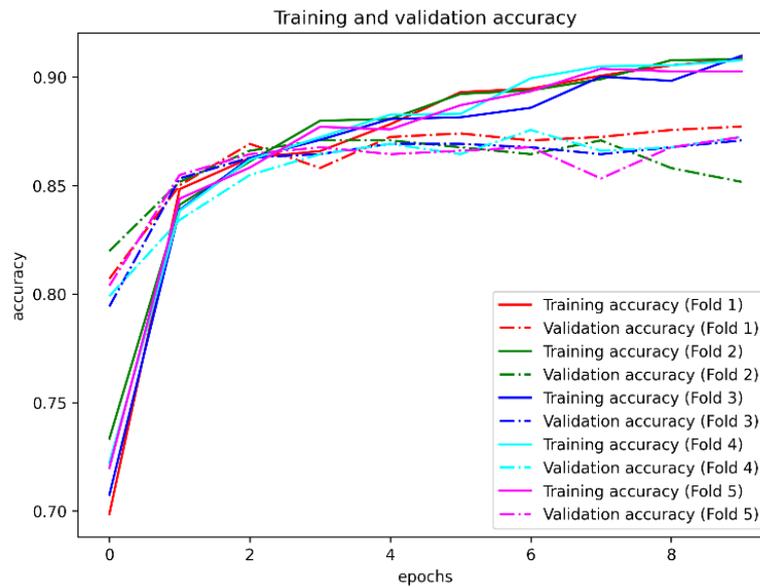


Figure 5. Accuracy Plot of Sentiment Prediction Model

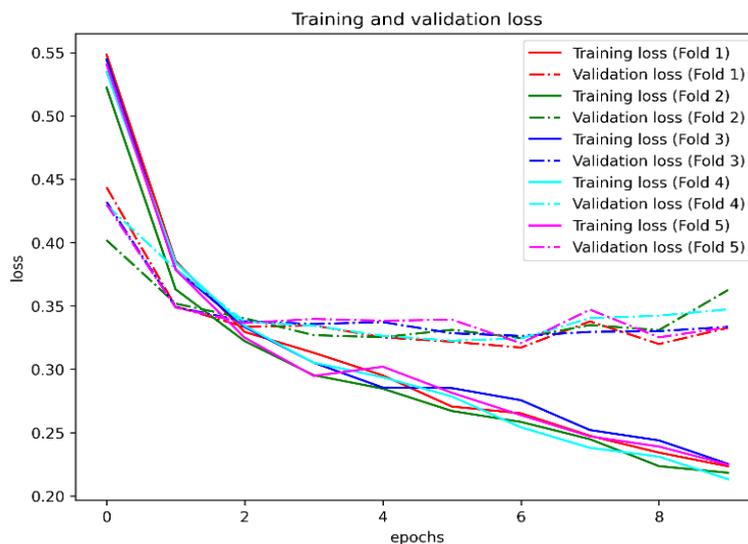


Figure 6. Loss Plot of Sentiment Prediction Model

Figure 5 and Figure 6 shows the training and validation accuracy and loss of each fold on K fold. The accuracy is increasing with increase of epoch and the loss is decreasing.

Similarly, KFold Cross-validation is used to evaluate the best target category prediction model using K=5. In Each Fold, data is split into training and testing sets with an 80:20 ratio. The category prediction model is trained with POS removed review with respect to category. KFold Cross-validation is used to evaluate the best category prediction model using K=5. In Each Fold, data is split into training and testing sets with an 80:20 ratio. The dropout layer was added to remove the over-fitting in our model. Models were tested with different combinations of hidden layers, dropouts, epochs, and batch size. The final chosen combination results in the following output of Table 5.

Table 5. K Fold Cross Validation Result of Target Category Prediction Model

Fold	Loss	Accuracy	Validation Loss	Validation Accuracy
1	0.7167	0.7501	0.6548	0.7895
2	0.5504	0.8184	0.6423	0.7863
3	0.5677	0.8032	0.6384	0.7911
4	0.5542	0.8072	0.6276	0.7927
5	0.5890	0.8068	0.6373	0.7831

Table 5 Forth fold gave the best result having a minimum validation loss of 0.6276 and validation accuracy of 0.7927. So, this model was chosen for the target category classification task. The accuracy and loss are good in each fold which is showing the good performance of our model. Figure 7 and Figure 8 show the training and validation accuracy and loss of each fold on the K fold. The accuracy is increasing with the increase of epoch and the loss is decreasing.

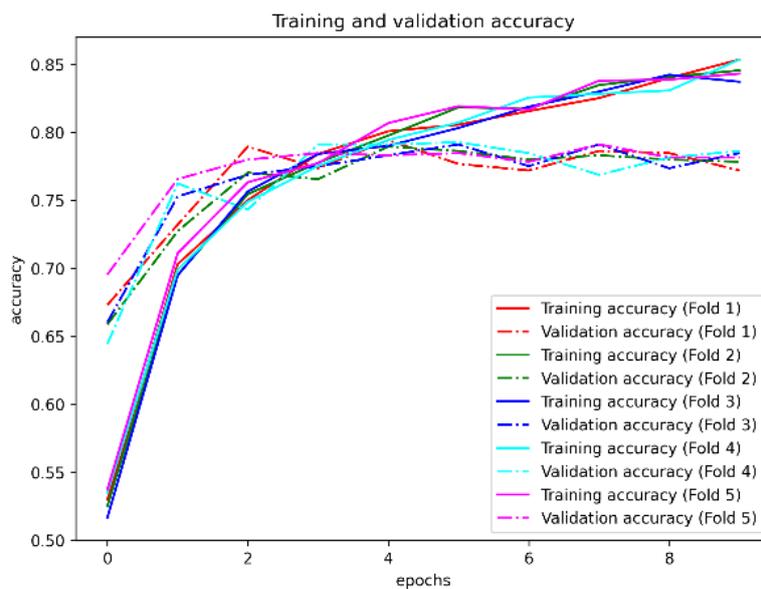


Figure 7. Accuracy Plot of Target Category Prediction Model

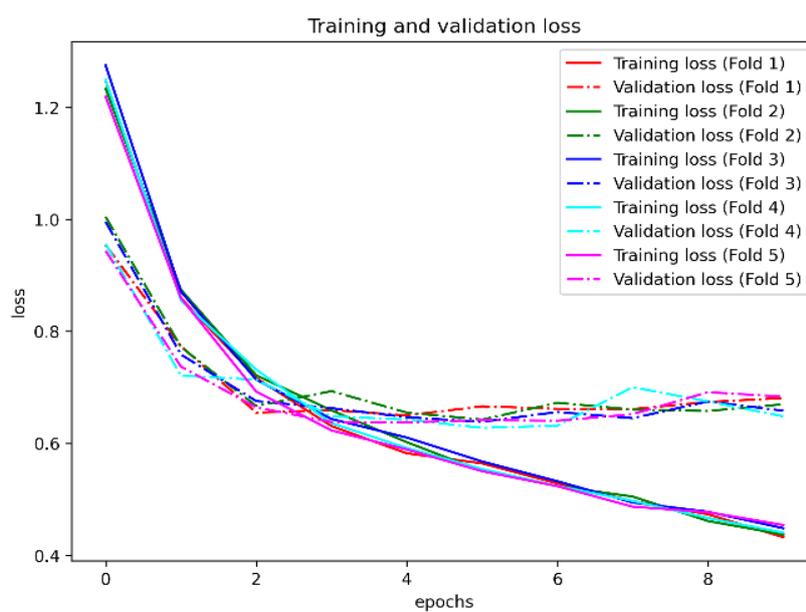


Figure 8. Loss Plot of Target Category Prediction Model

4.2 Validation

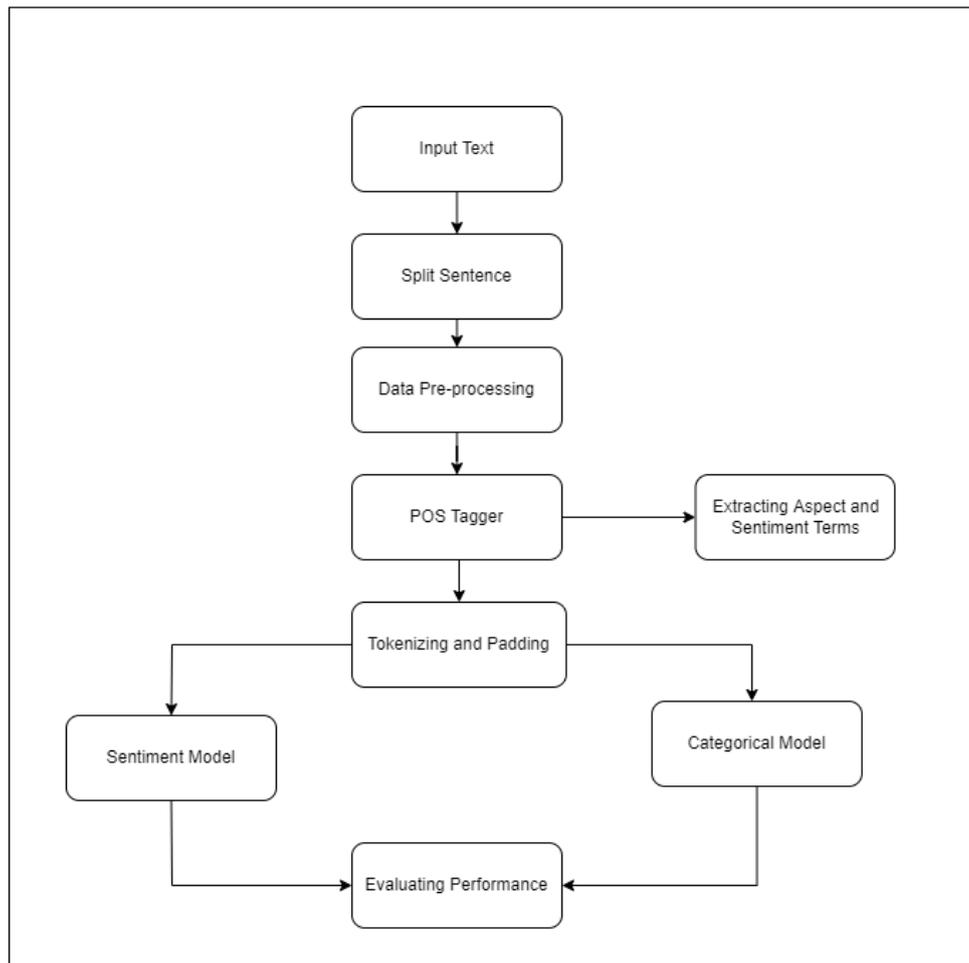


Figure 9. Validation Flowchart

For the validation purpose new validation dataset is applied to validation flowchart in Figure 9. The input sentence is split to different sentence by using conjunctions like **र**, **तर**, **अनि**. Data Pre-processing and POS tagging is done, sentiments and aspect terms are extracted from POS tagging. Data is tokenized padded and then passed to saved sentiment and categorical model. The performance of two models is evaluated using confusion matrix, accuracy, precision, recall and F1 Score. Figure 10 shows the confusion matrix for the sentiment prediction model. It is predicting well. Figure 11 shows the classification report for the sentiment prediction model on validation data. The F1 score of the model is 0.88 which is a good result.

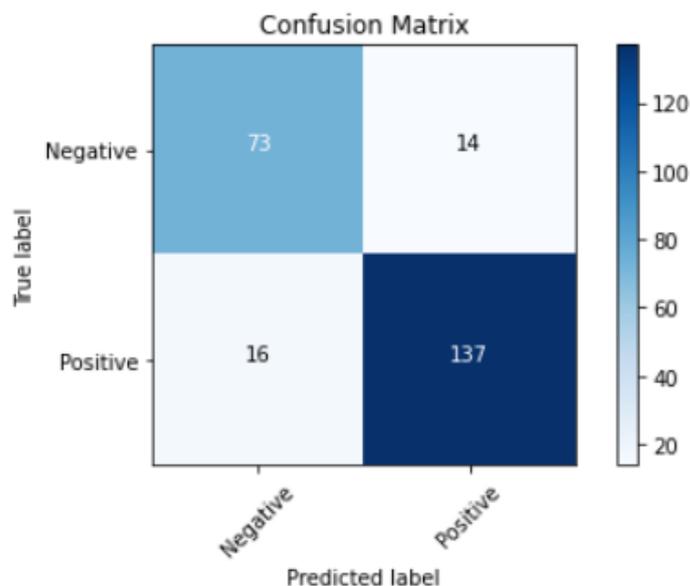


Figure 10. Confusion Matrix of Sentiment Prediction Model

	precision	recall	f1-score	support
0	0.82	0.84	0.83	87
1	0.91	0.90	0.90	153
accuracy			0.88	240
macro avg	0.86	0.87	0.87	240
weighted avg	0.88	0.88	0.88	240

Figure 11. Classification Report of Sentiment Prediction Model

Figure 12 shows the confusion matrix for the target category prediction model. It is predicting well. Figure 13 shows the classification report for the sentiment prediction model on validation data. The F1 score of the model is 0.76 which is a good result for multi class classification task.

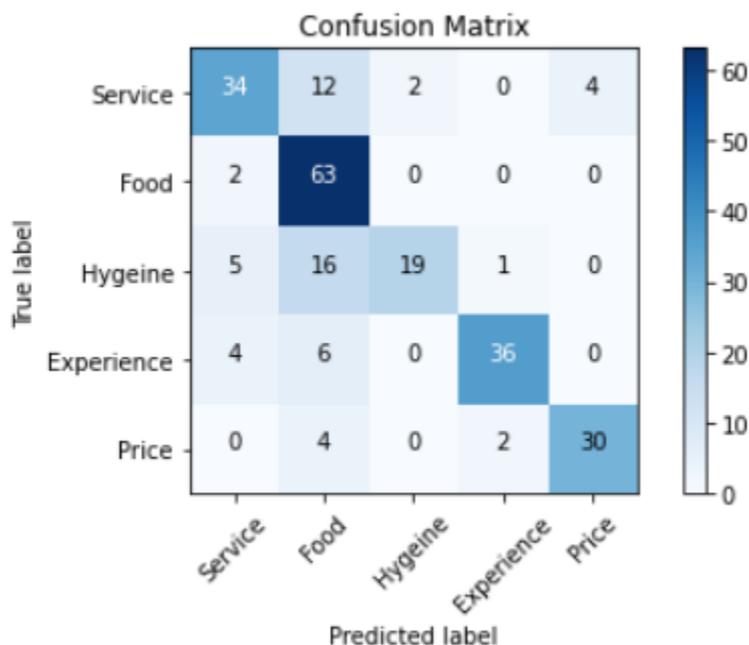


Figure 12. Confusion Matrix of Target Category Prediction Model

	precision	recall	f1-score	support
0	0.76	0.65	0.70	52
1	0.62	0.97	0.76	65
2	0.90	0.46	0.61	41
3	0.92	0.78	0.85	46
4	0.88	0.83	0.86	36
accuracy			0.76	240
macro avg	0.82	0.74	0.76	240
weighted avg	0.80	0.76	0.75	240

Figure 13. Classification Report of Target Category Prediction Model

The final aspect based output is predicted as follow

Input : खाना मिठो थियो तर पानी फोहोर थियो

Output :

Sentiment is मिठो , Aspect is खाना , Sentiment is ‘Positive’, Category is Food

Sentiment is फोहोर, Aspect is पानी, Sentiment is ‘Negative’, Category is Hygiene

Figure 14. Aspect Based Final Output

5. Conclusion

This work focuses on the creation of a new aspect sentiment analysis dataset and a deep learning BiLSTM model for restaurant reviews in the Nepali language. The model for Sentiment classification classifies data into Positive and Negative. The model for category classifications classifies data into 5 different target terms Food (F), Price (P), Service (S), Hygiene (H), and Experience (E). We can classify data into aspect-based sentiment classes using these two models. We found the good performance of models in classifying polarities and categories. The training and validation accuracy for the polarity prediction model is 0.8830 and 0.8644. The training and validation accuracy for the category prediction model is 0.8072 and 0.7927. The F1 score for the new validation dataset for polarity and category prediction model is 0.88 and 0.76 respectively. The performance of a model in predicting each category is good.

To improve the task further, improvement on the quality of data can be done. We can further make a model more robust by training with more data. We focused on restaurant review data for our work. A more general-purpose model can be prepared by training with more diverse data. Further work on POS tagger can be done. Work on improving the accuracy of POS tagger can be done by using different rule-based tagging methods or probabilistic methods. The approach used in our work works well for a simple sentence. Research on other approaches can be done to handle complex sentences.

Acknowledgement

We'd like to express our special thanks of gratitude to the Department of Electronics and Computer Engineering, Pulchowk Campus for providing continuous support throughout the research. We would also like to thank all teachers and colleagues for their direct and indirect help related to research.

References

- [1] Akhtar, Md Shad, Asif Ekbal, and Pushpak Bhattacharyya. "Aspect based sentiment analysis in Hindi: resource creation and evaluation." In Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16), pp. 2703-2709. 2016.
- [2] Bal, Bal Krishna, and Prajol Shrestha. "A morphological analyzer and a stemmer for Nepali." PAN Localization, working papers 2007 (2004): 324-31.

- [3] Tamrakar, Sujana, Bal Krishna Bal, and Rajendra Bahadur Thapa. "Aspect based sentiment analysis of Nepali text using support vector machine and naive bayes." *Technical Journal* 2, no. 1 (2020): 22-29
- [4] Singh, Oyesh Mann, Sandesh Timilsina, Bal Krishna Bal, and Anupam Joshi. "Aspect based abusive sentiment detection in Nepali social media texts." In *2020 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pp. 301-308. IEEE, 2020.
- [5] Loper, Edward, and Steven Bird. "Nltk: The natural language toolkit." *arXiv preprint cs/0205028* (2002).
- [6] Koirala, Pravesh, and Nobal B. Niraula. "Npvec1: Word embeddings for nepali-construction and evaluation." In *Proceedings of the 6th Workshop on Representation Learning for NLP (RepL4NLP-2021)*, pp. 174-184. 2021

Author's biography

Piyush Pant currently pursuing his MSc Degree in Computer System and Knowledge Engineering from Pulchowk Campus, Institute of Engineering, Tribhuvan University.

Subarna Shakya received MSc and PhD degrees in Computer Engineering from the Lviv Polytechnic National University, Ukraine, in 1996 and 2000, respectively. Currently, he is the Professor at the Department of Electronics and Computer Engineering, Pulchowk Campus, Institute of Engineering, Tribhuvan University. His research interest includes e-government system, distributed system, cloud computing, and software engineering and information system, Deep Learning, and Data Science.