Khulna University of Engineering & Technology

Course No: CSE 4120
Course Title: Technical Writing and Seminar

Bangla Aspect-Based Sentiment Analysis

Submitted By:
Irhamul Islam
Roll Number: 1907093
Department of Computer Science and Engineering
Khulna University of Engineering & Technology

Course Teachers:

Dr. K. M. Azharul Hasan
Professor
Department of Computer Science and Engineering
Khulna University of Engineering & Technology

Sunanda Das
Assistant Professor
Department of Computer Science and Engineering
Khulna University of Engineering & Technology

Abstract

Sentiment analysis has been a very interesting topic for researchers as it is very useful in analyzing social media comments, online product review comments, etc. The usual sentiment analysis determines the emotion behind a text. Aspect-based analysis gives opinions based on different aspects of the content. At first, it extracts aspects from the sentence and then separates them based on their sentiment. In this study, I have reviewed three research papers that worked on aspect-based analysis for the Bengali language. They worked on the same dataset using different approaches. Solving this problem will be very helpful in various aspects and will create more research scope in the associated field.

Contents

List of Tables

List of Figures

1 Introduction

Number of users of social media is increasing day by day. Billions of people are writing posts, and comments and giving views about something. In recent years, people have become more involved to online shopping. During any pandemic situation like COVID-19, engagement in online shopping sites was heavy. The customers are constantly giving reviews or complaints about the product or service they're taking. Sentiment Analysis can analyze texts and give predictions of these large number of reviews, and comments.Thus giving narrow insight into whether the statement is positive, negative or neutral.So,sentiment analysis has become an important topic for researchers.

Sentiment analysis labels the overall sentiment of a text. Whereas Aspect-Based Sentiment Analysis(ABSA) predicts the sentiment by detecting the aspects of a sentence.For example-

"Bangladesh is playing well."

For the above sentence, ABSA will predict the polarity positive under the aspect of sports.In ABSA, there are two steps - 1. term extraction and 2. sentiment classification means assigning the polarity.

For the Bengali language, research on sentiment analysis is ongoing and improving gradually. There is a lack of research on topics like ABSA in the Bengali language. There are some shortcomings like - a lack of accurate or annotated datasets.

The reviewed research papers tried to improve the accuracy of aspect-based analysis(ABSA) in the Bengali language.In the first paper, a new approach called PSPWA(priority Sentence Part Weight Assignment) is introduced by Forhad[12]. With CNN, this shows better results. In the second paper, they found that less pre-processing of data increases the accuracy[8]. Both of the papers worked on the publicly available dataset. In the third paper, proposed a new version of the available dataset and named it BAN-ABSA[1]. This dataset is annotated with aspects and corresponding sentiment. They also claimed that using Bi-LSTM performs better than CNN.

ABSA is useful in many scenarios like social media comments or posts, online shopping reviews, or newspaper comments. At a glance, it can tell us the portion of positive or negative reviews from the large dataset. Thus bringing efficient analysis in the related field.

2 Background/Problem Statement

Feedback from customers from social media, online reviews, newspaper comments, or customer online surveys is creating a large data every day. It is very painstaking and tough for humans to read all the content and understand the people's opinions from it.

Unlike sentiment analysis,which gives an overall prediction, Aspect-Based analysis gives a far more better understanding.For example-

"Food is delicious but the service is poor."

It can be seen clearly that this comment of a restaurant review indicates positive feedback for the food but a negative opinion of the service in the same sentence. In usual overall prediction, it will show the opinion is bad of a customer. But this neglects the fact that the customer was happy with the food. So, there are two aspects-food and service. Based on that, if sentiment is analyzed, that will give a better understanding. This is Aspect-Based Sentiment Analysis in short ABSA.

There are four sub-tasks for ABSA mentiond in SemEval[16]-

Aspect Term extraction

Detecting a list of particular aspect terms from a sentence.

Aspect Term Polarity

This assigns a label like positive, negative, or neutral in various aspect terms that have been extracted.

Aspect Category Detection

There are some predefined sets of aspect categories. Comparing with that, identifying the aspect categories of the given sentence.

Aspect Category Polarity

After detecting the category, determines the polarity for each aspect.

3 Review of Literature

ABSA in the Bangla language is first introduced in 2018,done by Rahman and Dey[19].The dataset is made publicly available.Liu,B. first introduced ABSA.He first described it's procedures and sub-tasks.[11].
For opinion expression extraction,Cahyadi built an ABSA model using Bi-LSTM and Convolutional Neural Network(CNN) for sentiment polarity[6].Nandan proposed that lexicon-based presents less performance than machine learning techniques[13].Nazir done a research to find out the challenges of ABSA[14]. Based on stacked auto-encoders boidini developed a ABSA model that achieved better f1-score than Rahman and Dey's research of 2019[4].Based on Recurrent Neural Network(RNN), Wahid did a study of sentiment analysis in Bangla text[24].
In SemEval, ABSA's task is divided into four subtasks[16].In aspect extraction,a rule-based approach has been propsed from product reviews[17].In a reserach,the authors introduced two new corpora in Czech language to solve ABSA[22].
To implement ABSA,clause level classification can give significant performance for both tasks[23].Among many created datasets for ABSA,one is named Sentihood containing 5215 sentences collected from yahoo[20].In Hindi,a dataset for sentiment analysis is created containing 5412 comments of Hindi product reviews[2].

4 Methodology

4.1 Bangla Aspect-Based Sentiment Analysis Based on Corresponding Term Extraction

4.1.1 Dataset Collection

Dataset used in this research is a publicly available dataset first implemented in ABSA by Rahman and Dey[19].Cricket and Restaurant are two datasets of Bangla language .These generally contains the feedback of people on this topics.Link for the dataset-`https://github.com/atik-05/Bangla_ABSA_Datasets`

Table 1: Overall summary of datasets

	Cricket	**Restaurant**
No. of Sentence	2979	2059
Aspect Category	Batting, Bowling, Team Management, Team, Other	Food, Service, Price, Ambiance, Miscellaneous
Aspect Polarity	Positive (19%) Negative (72%) Neutral (9%)	Positive (59%) Negative (23%) Neutral (6%) Conflict (12%)

4.1.2 Data Preprocessing

To make the dataset useable for the model, in pre processing step, unnecessary characters are removed. Here, punctuation and numerical words such as Dari(' |'), Comma(','), etc are eradicated as these carry no significant impact on the result.

4.1.3 Feature Enginneering

For better performance of any machine learning model, feature engineering plays an effective role. Using Term Frequency and Inverse Document Frequency(TF-IDF), the dataset has been tokenized.
Constant features do not affect target outputs. These values are negligible. That's why, these features are removed from the numbered matrix.
Correlated features are influenced by common mechanisms. Highly correlated features give redundant information. These values were also removed from the numbered matrix.
For removing constant features and correlated features, the threshold values were 0.005 and 0.75 respectively.

4.1.4 Priority Sentence Part Weight Assignment(PSPWA)

NOUNS of a sentence is focused on extracting aspect terms. It is considered that NOUNS are priority words.In any sentence, the portion holding a NOUN is the priority sentence part.

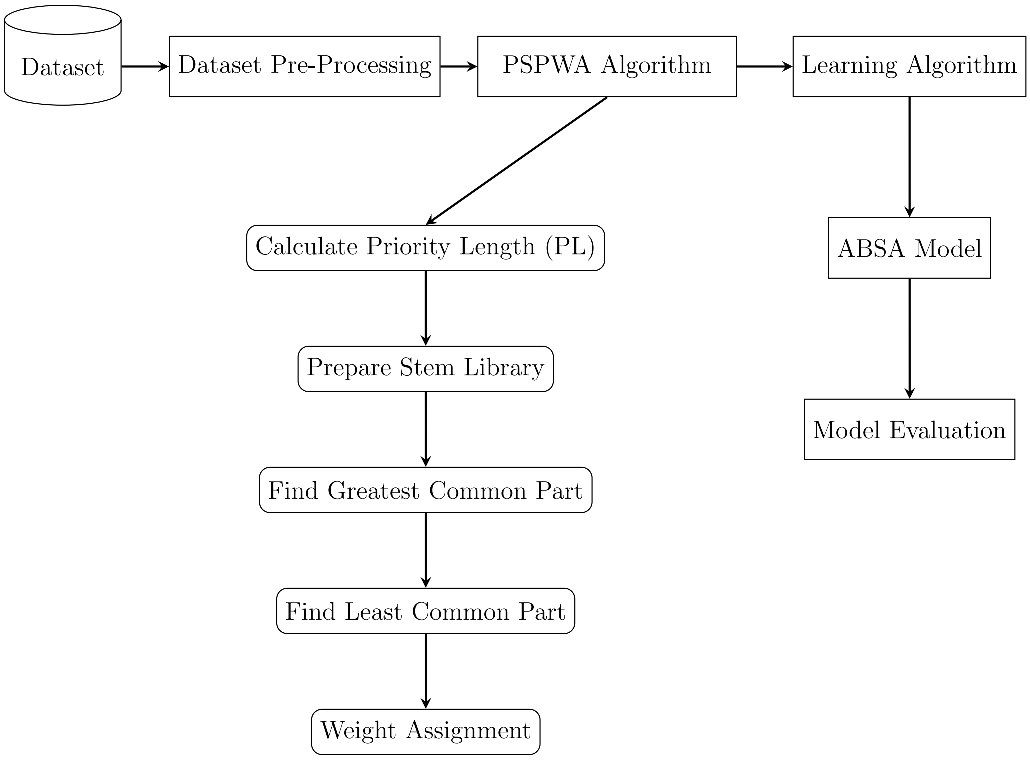

Figure 1: Diagram for proposed method for first paper

1. **Priority Length (PL)**:
   To find out the priority part, an equation is applied. This returns an integer value. The equation looks like below.

$$\mathrm{PL} = \min(\mathrm{LEN(SENTENCE)}, 3) + \frac{\mathrm{LEN(SENTENCE)}}{3} - 1 \tag{1}$$

2. **Stem/ Root Form of Verb**:
   Roots are the base from of the words.A suffix or prefix can be added to that word but expresses the same meaning.

3. **Greatest Common Part & Least Common Part**
   The greatest Common Part is considered to be the largest frequency sentence part of the stem. It is the priority part and the Least Common Part is the least frequency sentence part of the stem. It is not a priority part.

4. **Weight Assignment**:
   - Initially,weight is zero for each word.
   - Convert the words of sentence to stem.
   - If any word is present within the priority part, increase one
   - If sentence part of word matches with GCP of stem, increase one
   - If sentence part of word and GCP both are priority part, increase one
   - If sentence part of word don't match with LCP, increase one
   - If weight of word is greater than zero , then the weight is being multiplied with corresponding value of the word in TF-IDF matrix.
   - Then the result is replaced with the corresponding value in TF-IDF matrix. The equation looks like this-

$$Matrix_{i,j} = Matrix_{i,j} \times W_{i,j} \quad ; \quad \text{if } W_{i,j} > 0 W_{i,j} > 0 \tag{2}$$

5. **Normalization**:
   Min-Nax classifier is applied as a normalization technique to transform all features in a suitable range.Equation of Min-Max classifier is given below -

$$\text{Matrix}_{i,j,\text{scaled}} = \frac{\text{Matrix}_{i,j} - \text{MIN}(\text{Matrix}_{i,j})}{\text{MAX}(\text{Matrix}_{i,j}) - \text{MIN}(\text{Matrix}_{i,j})} \tag{3}$$

4.1.5 Algorithm analysis

Several learning algorithms have been used to implement the proposed method. Methods like K-nearest neighbor (KNN), Support Vector Machine (SVM), Linear Regression (LR), Random Forest (RF), and Convolutional Neural Network (CNN) have been employed. The values of the parameters are given below. These values are determined by trying many experimental values.

CNN Architecture:

Two steps - 1.Feature extraction and 2.Classification

- Feature extraction is a combination of convolutional and pooling layers.
- The first CONV layer identifies low-level features.It becomes adaptable to high-level features if more layers are added.

Table 2: Parameter list of CNN and supervised learning algorithms

Classifier	**Parameter**
KNN	No of neighbors = 3, Metric = Euclidean
SVM	Max iterations = 1000, Multi class = 'ovr'
LR	Solver = 'sag', Max iterations = 150
RF	Criterion = 'gini', No of estimators = 150
CNN	Optimizer = 'adam', Learning rate = 0.0005, Batch Size = 100, Epochs = 20, Convolution Activation = 'relu', Dense Activation = 'softmax', Loss Function = 'crossentropy'

- The Classification step, based on information on CONV layers, classifies output.
- Learning rate of 0.0005 using ADAM as it converges faster[7]
- Cross-entropy has been used as a loss function
- RELU used as a convolutional activation function
- Proposed CNN has seven layers.
- The output of the TF-IDF numbered matrix will be input for the CONV layer. Here, padding remains the same.

The dataset is divided into two sets test data and train data. Giving 80% of the dataset to the training dataset and 20% to the testing dataset.

4.2 Aspect Based Sentiment Analysis in Bangla Dataset Based on Aspect Term Extraction

4.2.1 Dataset Collection

The publicly available dataset for ABSA by Rahman and Dey of 2018 is used for analysis[19]. This dataset is created for the first time in Bengali.Link for the dataset-`https://github.com/atik-05/Bangla_ABSA_Datasets`
In the Cricket dataset, humans annotated the comments in five different aspects -bowling, batting, team, team management, and others. The restaurant dataset is translated into Bengali from the SemEval English dataset[16]. Here also, five aspects-Food, Price, Service, Ambiance, and Miscellaneous.The dataset overview can be seen in Table 1.

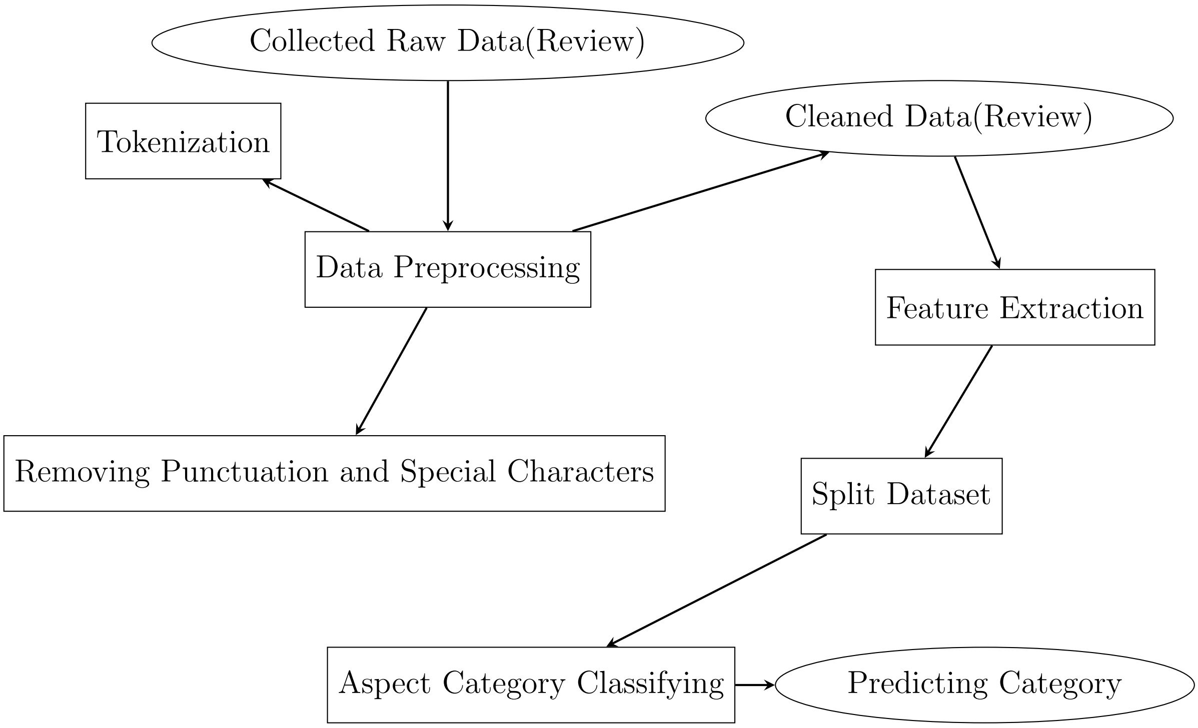

Figure 2: Proposed Model of Corresponding Term Extraction

4.2.2 Data Preprocessing

Data preprocessing is a significant step in doing natural language processing tasks. The given dataset may contain some redundant characters that don't add any value to the final output or it may contain some information that may add complexity for the model analysis.

Here,text data is reduced by various steps which made the input size smaller.The steps are -

1. **Removing special characters**:
   Special characters lead to ambiguity in the model often. That's why, they are removed from the dataset. Otherwise, unnecessary complexity would have been created.

2. **Removing punctuations**:
   Removing punctuations is a very well-known preprocessing step.As there is change even in punctuation from language to language, Punctuations were removed.

4.2.3 Feature Extraction

In this procedure, texts have been converted to numeric form to use them as a feature. The procedure follows –

1. **BOW**:
   To train any statistical algorithm using machine learning, the dataset has to be in numeric form.To ensure the functionality, the texts are converted to numbers.Bag Of Words can be used as a technique to implement this. This reflects the frequency of words in a textual content[5].

2. **TF-IDF** :
   It is evident from the name that it contains two words –TF and IDF. The first one means Term Frequency and the latter means Inverse Document Frequency.
   TF counts the frequency of a word in a content.IDF indicates the importance and weight of that word..

Equation for TF and IDF is given below:

$$\mathrm{TF}(t) = \frac{\text{Number of times term } t \text{ appears in a document}}{\text{Total number of terms in the document}} \tag{4}$$

$$\mathrm{IDF}(t) = \log_e \left(\frac{\text{Total number of documents}}{\text{Number of documents with term } t \text{ in it}} \right) \tag{5}$$

Sklearn library[15] is used which contains the TFiDfVectorizer class. This is used to convert the feature into TF-IDF feature vectors. limitation of maximum features was set to 2500.

4.2.4 Algorithm Analysis

Several trials and errors have to be faced before finalizing which machine learning algorithm will be used to determine the outcome. This is a very rigorous process as the algorithm efficiency varies from dataset to dataset. The algorithms used in previous research have also been considered. The algorithms mentioned below are selected finally to classify.

1. To compare with the result of Rahman & Dey[19]:

   - SVM (Support Vector Machine)

    - RF (Random Forest)
    - KNN (K-Nearest Neighbor)

2. Additionally used:

    - LR (Logistic Regression) [10]
    - NB (Naïve Bayes)

Logistic Regression works best for prediction and classification problems. It is a suitable regression analysis to work on binary variables. To find the relationship between one dependent variable and several independent variables, LR was used here.
Naïve Bayes has good efficiency for predictive modeling. It usually assumes that every input is independent. NB is selected as it is less affected by data scarcity. It is often used in sentiment analysis tasks.
The dataset is divided as 80% for training and 20% for testing.

4.2.5 Languages and Tools

- Python 3 (Jupyter Notebook) in Anaconda
- Python modules: scikit-learn[15]
- NLTK (A leading platform for building Python programs)[3]

4.3 BAN-ABSA: An Aspect-Based Sentiment Analysis dataset for Bengali and it's baseline evaluation

4.3.1 Data Collection

There has always been a lack of a proper dataset for the Bengali language in the field of aspect-based sentiment analysis. In past research, two datasets were present but they contained very few comments.

Here, a new high-standard dataset is created.it is named BAN-ABSA. has 9009 comments from many Bangali news sites.The dataset has sentences containing :

- Aspects:
    - Politics

  - Sports
  - Religion
  - Others

- Polarity:
  - Positive
  - Negative
  - Neutral

Comments were collected from several news portals where people expressed their opinions and views about any particular matter. This could be comments or posts. The popular news sites were selected as they had more engagement of people.Data collection overview:

- Daily Prothom Alo: 16735 comments
- Daily Jugantor: 14389 comments
- Kaler Kantho: 8062 comments

These 39186 comments were collected.

4.3.2 Annotating Data

After collecting the comments, they were preprocessed. Multi-lined comments have been removed or multi-lined comments trimmed into a singe-line comments. Emojis were also removed.

In the annotating phase, the data was divided into three parts. They were distributed among 9 annotators. Each comment was annotated by 3 annotators. They analyzed the data for aspect and polarity. The comments can be in one of the three aspects: politics, religion, and sports. If the comment that doesn't match with the mentioned aspects, goes to the other aspect. The polarity can be either positive, negative, or neutral. Majority voting was followed to avoid any contradiction.

For example, to do aspect annotation, say two annotators marked a sentence as politics, and one annotator marked it as Religion aspect. From majority voting, it was decided to annotate this comment as having a political aspect. No tie situation was faced.

Finally, 9009 comments were labeled. The dataset contains four aspects and 3 polarities. The dataset was created such it maintains balance. Like, in the case of news, people write no comments for good news. On opposite, a lot of comments can be found on a post about any negative incident.

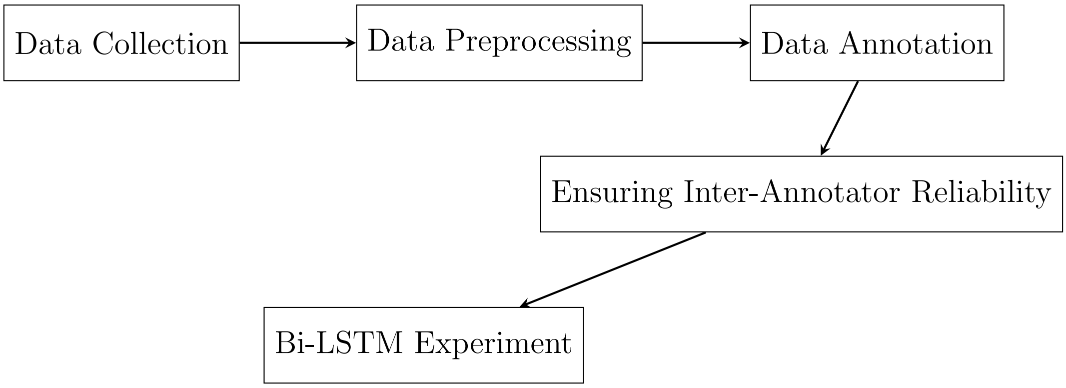

Figure 3: Proposed Model for BAN-ABSA

4.3.3 Dataset Analysis

To ensure annotator's reliability, Intra-class Correlation Coefficient (ICC) was calculated.It came 0.77.
Zipf's law was also applied [18] in our dataset.The law states that, the collection frequency (cfi) of the ith most common term in the dataset should be proportional to 1/i .

$$c_{f_i} \propto \frac{1}{i} \tag{6}$$

4.3.4 Experiment With Bi-Directional Long Short Term Memory

After creating the benchmark dataset, BAN-ABSA, which can be used for both the sub-tasks: aspect extraction and sentiment classification.
In the pre-processing phase, The comments were removed containing English words. All the punctuation marks were removed. Then, the data was tokenized.
Some deep neural network models and some traditional supervised machine learning models were used to complete this task. In the experiment, Bi-LSTM showed good performance achieving the highest f1-score.

- **Bi-Directional Long Short Term Memory**
  Recurrent Neural Networks (RNN) show good results and they are highly used in text classification.But, RNN has a vanishing gradient problem.To resolve this issue, LSTM was proposed [9].
  In LSTM, information can only be transferred in forward states. At any time, the state is dependent only on past information. But forward information may also be required in some cases.
  BiLSTM is introduced to solve this problem. BiLSTM's architec-

ture consists of two hidden, opposite-direction LSTM layers to completely hold the input context.Regular RNN's state neurons are split into two parts.One part captures features of the past and uses forward states.The other one takes features of the opposite direction and uses backward states[21].

5 Result Analysis

F1-score as an evaluation matrix is presented to compare the three papers for result accuracy and to evaluate model performance. The papers used various learning algorithms. Here, not all algorithms are followed the same in all papers. Like, Forhad[12] used CNN but Haque et al.[8] did not.
The True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) are four parameters to measure the terms f1-score, precision, and recall.

$$\text{Precision} = \frac{TP}{TP + FP} \tag{7}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{8}$$

$$\text{F1-score} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{9}$$

In the following table,the first two papers are compared .Both Forhad[12] and Haque et al.[8]worked on the same dataset.

Method	**Dataset**	**Paper 1**	**Paper 2**
SVM	Cricket	0.48	0.35
KNN	Cricket	0.31	0.27
RF	Cricket	0.41	0.37
LR	Cricket	0.48	0.34
NB	Cricket	0.39	0.18
CNN	Cricket	0.59	not applied

Table 3: F1-Score comparison on cricket dataset

It can be seen that Forhad[12] performs better than Haque et al.[8] significantly using the method SVM and LR.
These two papers also worked on the restaurant dataset.Table 4 shows the comparison.

Method	Dataset	Paper 1	Paper 2
SVM	Restaurant	0.52	0.39
KNN	Restaurant	0.39	0.38
RF	Restaurant	0.35	0.35
LR	Restaurant	0.44	0.43
NB	Restaurant	0.34	0.17
CNN	Restaurant	0.67	not applied

Table 4: F1-Score comparison on restaurant dataset

Here also , Forhad[12] performed better than Haque et al.[8].
The third paper created a new benchmark dataset named BAN-ABSA[1].They worked on that showing a significant good result.The following table shows the f1-score for the corresponding algorithms. This paper shows a far better

Method	Dataset	Paper 3
SVM	BAN-ABSA	0.69
KNN	BAN-ABSA	0.47
RF	BAN-ABSA	0.65
CNN	BAN-ABSA	0.75

Table 5: F-1 score table for BAN-ABSA

f1-score than the rest two in aspect-based sentiment analysis.Apart from the methods mentioned in the table, they also used LSTM and Bi-LSTM.
Using LSTM and Bi-LSTM for aspect extraction on the proposed dataset, they achieved f1-score of 77.24 and 79.38 respectively.

6 Findings and Recommendations

From this rigorous study of three research papers solving the same problem, some insights have been gained.They are mentioned below:

- Feature engineering reduces dataset shape.This phase excludes the constant and correlated features.The refined dataset is more suitable for the learning model.
- The PSPWA technique mentioned in the first paper by Forhad[12] replaced the TD-IDF score of the numbered matrix by assigning a new weight.

- Overfitting is an issue for tasks like sentiment analysis. This problem can be solved by adding more training samples or reducing features.
  In the feature engineering phase, the dataset is already reduced by eliminating some redundant data. As the first two papers,Forhad[12],Haque et al.[8] used publicly available datasets, adding more samples could not be implemented.
  The third paper,BAN-ABSA[1] created its own high standard dataset which showed better result

- In the first paper,Forhad[12], applying PSPWA in the data pre-processing step has helped to gain a better F1-Score than previous studies.

- The second paper,Haque et al.[8] has provided better results by reducing some preprocessing steps. This less preprocessing leads to better F1 scores in both datasets (Restaurant and Cricket).
  Aspect category extraction is a multi-label classification problem [5] where one sentence might carry several categories.That's why, the proposed method here may skip some aspect categories.There is room for improvements.
  One solution can be by training the dataset using POS tagging.

- As for the learning algorithms, several algorithms have been used to evaluate the proposed model in all three papers. In general, it can be said that CNN and LR performed better than the other methods.
  Additionally, for aspect term extraction LSTM and Bi-LSTM are used in the third paper,BAN-ABSA[1].This showed far better accuracy than other methods.

- The dataset plays a vital role in aspect-based sentiment analysis. Rather than the publicly available dataset, BAN-ABSA[1] showed far better performance.

After analyzing the findings by studying the three research papers, I highly recommend the third paper, **BAN-ABSA**[1] for the best performance among those. It showed good results both in terms of aspect extraction and polarity of sentence detection.I propose the Bi-LSTM architecture to be used in Aspect Based Sentiment Analysis. The main reasons are :

- Creating a high-standard dataset for the Bengali language.

- Achieving high accuracy for aspect extraction by using Bi-LSTM.

- Better f1-score than the remaining papers in sentiment analysis. .

7 Addressing Course Outcomes and Program Outcomes

The course titled "Technical Writing and Seminar" has outcomes in the following way:

- **Problem Analysis**:
  In this course, a problem was analyzed. To implement this, three research papers were selected.Each of them tried to solve the problem with different approaches.the papers were studied thoroughly and an understanding was developed of the methodology or the solution approach. By doing so, it is learned how to think of different solutions for a particular issue.

- **Ethics**:
  Throughout this course, ethical values were maintained. For example, this report, it was tried to maintain the highest level of integrity which means - not copying other works or giving credit to someone whose work has been borrowed.

- **Individual and Teamwork**:
  The task of problem analysis was done individually. By analyzing three papers, it is learned that how to study research papers and make a review of those papers individually.
  Throughout the course, in many cases, teamwork was involved. In regular sessionals, tasks were given. Sometimes, they were solved collaboratively with the help of others. Thus, ensuring teamwork ability.

- **Communication**:
  Communication skill was developed by giving effective presentations, communicating with course teachers regarding engineering activities, etc. Designing documents or reports also helped to furnish one's communication ability.

8 Addressing Complex Engineering Activities

The following attributes of Complex Engineering Activities was gained:

- **Range of resources**:
  In analyzing a problem, resources can vary for a large range. Informa-

tion is needed to check the validity of the research papers. To understand the studies, other knowledge was gained from various sources. This report is written using latex which is also a resource.

- **Level of interaction**:
  While understanding the problem, interaction to different people or between different domains of topics happened.

- **Innovation**:
  The three research papers tried to solve the problem in their unique way. Thus, leading to an innovation.

- **Consequences for society and the environment**:
  Solving this problem will be impactful for society,Sentiment analysis has various applications in social media monitoring in online shopping reviews, etc.

- **Familiarity**:
  There were some research studies done on this topic before. The papers mentioned tried to improve the solution by proposing a new methodology.Although there are familiarity, each of the papers contributes in a different way.

9 Conclusion

In this study, three research papers were discussed solving the problem of aspect-based sentiment analysis. Each paper proposed different approaches to solve this issue.
The ABSA model works into two parts - 1. aspect terms extraction and 2. aspect polarity determination.A new technique named PSPWA is introduced for extracting aspect terms which has improved the F1-score.
With less preprocessing, gives better outputs for aspect extraction. Two traditional steps can be used to clean data.With a proper dataset, this result can be improved.The BAN-ABSA dataset[1] is a significant contribution in research in this field.
Resource barriers hindered many researchers from doing work on Bengali ABSA. BAN-ABSA[1] can be used in further ABSA tasks and create more research scope.
Sentiment analysis is useable for various cases like spam review or comment detection and fraud app detection. This can help to make better business decisions also improve cyber security.

In the future, more advanced techniques of deep learning can be applied in NLP for ABSA. Continuous exploration may lead to finding better learning techniques. Also, the Bangla POS tagger can be explored to train the model in aspect term extraction and train classifier.
The found results are not too high compared to English ABSA. Modifications need to be made to improve the overall performance. In the Bengali language, this type of research needs to be done more as there is scope for improvements.

References

[1] Mahfuz Ahmed Masum, Sheikh Junayed Ahmed, Ayesha Tasnim, and Md Saiful Islam. Ban-absa: An aspect-based sentiment analysis dataset for bengali and its baseline evaluation. In *Proceedings of International Joint Conference on Advances in Computational Intelligence: IJCACI 2020*, pages 385–395. Springer, 2021.

[2] Md Shad Akhtar, 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)*, pages 2703–2709, 2016.

[3] Steven Bird, Ewan Klein, and Edward Loper. *Natural language processing with Python: analyzing text with the natural language toolkit.* " O'Reilly Media, Inc.", 2009.

[4] Matteo Bodini. Aspect extraction from bangla reviews through stacked auto-encoders. *Data*, 4(3):121, 2019.

[5] Jason Brownlee. A gentle introduction to the bag-of-words model, aug. 2019. *URL https://machinelearningmastery. com/gentle-introduction-bag-words-model.*

[6] Alson Cahyadi and Masayu Leylia Khodra. Aspect-based sentiment analysis using convolutional neural network and bidirectional long short-term memory. In *2018 5th international conference on advanced informatics: concept theory and applications (ICAICTA)*, pages 124–129. IEEE, 2018.

[7] M Hamed, A El Desouky, et al. Effect of learning rate on the recognition of images. *Active and passive electronic components*, 19:1–12, 1996.

[8] Sabrina Haque, Tasnim Rahman, Asif Khan Shakir, Md Shohel Arman, Khalid Been Badruzzaman Biplob, Farhan Anan Himu, Dipta Das, and Md Shariful Islam. Aspect based sentiment analysis in bangla dataset based on aspect term extraction. In *Cyber Security and Computer Science: Second EAI International Conference, ICONCS 2020, Dhaka, Bangladesh, February 15-16, 2020, Proceedings 2*, pages 403–413. Springer, 2020.

[9] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

[10] Murat Korkmaz, SELAMİ Güney, and ŞULE Yiğiter. The importance of logistic regression implementations in the turkish livestock sector and logistic regression implementations/fields. *Harran Tarım ve Gıda Bilimleri Dergisi*, 16(2):25–36, 2012.

[11] Bing Liu. *Sentiment analysis and opinion mining*. Springer Nature, 2022.

[12] Forhad An Naim. Bangla aspect-based sentiment analysis based on corresponding term extraction. In *2021 International Conference on Information and Communication Technology for Sustainable Development (ICICT4SD)*, pages 65–69. IEEE, 2021.

[13] Neha Nandal, Jyoti Pruthi, and Amit Choudhary. Aspect based sentiment analysis approaches with mining of reviews: A comparative study. *International Journal of Recent Technology and Engineering*, 7(6):95–99, 2019.

[14] Ambreen Nazir, Yuan Rao, Lianwei Wu, and Ling Sun. Issues and challenges of aspect-based sentiment analysis: A comprehensive survey. *IEEE Transactions on Affective Computing*, 13(2):845–863, 2020.

[15] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. Scikit-learn: Machine learning in python. *Journal of machine learning research*, 12(Oct):2825–2830, 2011.

[16] Maria Pontiki, Dimitris Galanis, Haris Papageorgiou, Ion Androutsopoulos, Suresh Manandhar, Mohammed AL-Smadi, Mahmoud Al-Ayyoub, Yanyan Zhao, Bing Qin, Orphée De Clercq, et al. Semeval-2016

task 5: Aspect based sentiment analysis. In *ProWorkshop on Semantic Evaluation (SemEval-2016)*, pages 19–30. Association for Computational Linguistics, 2016.

[17] Soujanya Poria, Erik Cambria, Lun-Wei Ku, Chen Gui, and Alexander Gelbukh. A rule-based approach to aspect extraction from product reviews. In *Proceedings of the second workshop on natural language processing for social media (SocialNLP)*, pages 28–37, 2014.

[18] David MW Powers. Applications and explanations of zipf's law. In *New methods in language processing and computational natural language learning*, 1998.

[19] Md Atikur Rahman and Emon Kumar Dey. Datasets for aspect-based sentiment analysis in bangla and its baseline evaluation. *Data*, 3(2):15, 2018.

[20] Marzieh Saeidi, Guillaume Bouchard, Maria Liakata, and Sebastian Riedel. Sentihood: Targeted aspect based sentiment analysis dataset for urban neighbourhoods. *arXiv preprint arXiv:1610.03771*, 2016.

[21] Mike Schuster and Kuldip K Paliwal. Bidirectional recurrent neural networks. *IEEE transactions on Signal Processing*, 45(11):2673–2681, 1997.

[22] Ales Tamchyna, Ondrej Fiala, and Katerina Veselovská. Czech aspect-based sentiment analysis: A new dataset and preliminary results. In *ITAT*, pages 95–99, 2015.

[23] Tun Thura Thet, Jin-Cheon Na, and Christopher SG Khoo. Aspect-based sentiment analysis of movie reviews on discussion boards. *Journal of information science*, 36(6):823–848, 2010.

[24] Md Ferdous Wahid, Md Jahid Hasan, and Md Shahin Alom. Cricket sentiment analysis from bangla text using recurrent neural network with long short term memory model. In *2019 International Conference on Bangla Speech and Language Processing (ICBSLP)*, pages 1–4. IEEE, 2019.

SL No.	Title	Authors	Publishers	Source	Year
1	Bangla Aspect-Based Sentiment Analysis Based On Corresponding Term Extraction[12]	Forhad An Naim	IEEE	International Conference on ICT	2022
2	Aspect Based Sentiment Analysis in Bangla Dataset Based on Aspect Term Extraction[8]	Sabrina, Tasnim, Asif, Shohel, Biplod, Himu	Springer	Second EAI International Conference	2023
3	BAN-ABSA: An Aspect-Based Sentiment Analysis Dataset for Bengali and its Baseline Evaluation[1]	Masum, Junaed, Ayesha, Saiful	Springer	Proceedings of International Joint Conference on Advances in Computational Intelligence	2024

Table 6: Publication Details